

University of Naples “Federico II”



PhD Program

“Human Reproduction, Development and Growth”

Director

Prof. Claudio Pignata

PhD Thesis

***New strategies for the study  
of Celiac Disease genetics***

Candidate

Valentina Izzo

Tutor

Prof. Luigi Greco

Academic Year: 2012-2013

## **Abstract**

Celiac Disease (CD) is a chronic enteropathy affecting the small intestine and triggered by gluten proteins contained in wheat, barley and rye. It is characterized by an autoimmune response in genetically susceptible individuals, and only the 40% of the genetics is explained by HLA predisposition. Recently, several susceptibility loci not HLA related have been identified by genome-wide association studies (GWAs). Our aim is to extract as many information as we can from the large amount of data emerged from these studies.

Firstly, we tried to improve the analyses of candidate genes characterizing their expression profile and comparing the results obtained on mucosal biopsies, peripheral monocytes and T-cell lines isolated from duodenal mucosa, in order to better understand the variation in the celiac type and to correlate it with a model of CD pathogenesis. Furthermore we investigated the non-HLA genetics of Potential CD patients and this allowed us to speculate that these patients are a mixture of two populations, with different non-HLA genetics. We also used the new genetic findings to build a risk model for CD using HLA and non-HLA risk alleles. In both these two last cases we aimed to significantly improve the identification of subjects at risk to develop CD.

The critic analysis of the large panorama of CD candidate genes leads to consider several mechanisms of action that may have gone unnoticed if we persist in considering all the genes together. Whereas, analyzing a small set of genes from different points of view it is possible to really understand why this more than the other gene resulted to be associated with the disease. GWAs made the history of complex diseases, but represent only the starting point from which postulate hypothesis, mechanisms of action, interactions, in order to untangle the skein represented by the complex pathogenesis of celiac disease.

<b>Introduction</b>	<b>5</b>
<b>Celiac Disease</b>	<b>5</b>
Diagnosis	6
Genetic risk factors	7
HLA genes	7
Non-HLA genes	9
Function of selected candidate genes	15
<b>Aim</b>	<b>19</b>
<b>Chapter 1. Non-HLA genes and Celiac Disease: Gene expression studies</b>	<b>22</b>
<b>Background and aim</b>	<b>22</b>
<b>Materials and methods</b>	<b>24</b>
Patients	24
Monocytes	24
Gliadin preparation	25
Generation of gliadin-specific T-cell lines and functional assay	25
Expression studies	26
Statistical Analysis	27
<b>Results</b>	<b>27</b>
Expression studies	27
Mucosal tissue	27
Peripheral blood monocytes	30
T-cell lines isolated from duodenal biopsies.	33
Statistical Analyses	36
Discriminant Analysis in small intestinal tissue	36
Discriminant analysis in peripheral blood monocytes	38
<b>Discussion</b>	<b>41</b>
<b>Chapter 2. Non-HLA genes and Celiac Disease: Potential CD patients</b>	<b>45</b>
<b>Backgroun and aim</b>	<b>45</b>
<b>Methods</b>	<b>47</b>
Subjects	47
Genotyping	48
Statistical Analysis	49
<b>Results</b>	<b>49</b>
HLA typing	49
Eight candidate genes SNPs genotype and allele distribution	51
Five Candidate SNPs on the 4q genomic region	54
<b>Discussion</b>	<b>58</b>
<b>Chapter 3. Non-HLA genes and Celiac Disease: A Family study</b>	<b>62</b>
<b>Background and aim</b>	<b>62</b>
<b>Methods</b>	<b>63</b>
Family Samples	63
Non-HLA Single Nucleotide Polymorphisms (SNPs) typing	64

Analysis strategy and statistics	65
<b>Results</b>	<b>67</b>
SNPs evaluation and computation of the Bayesian Score	67
Validation of the BS and testing of a classification model	69
<b>Discussion</b>	<b>73</b>
<b><i>Conclusion and perspective</i></b>	<b>79</b>
<b><i>References</i></b>	<b>83</b>
<b><i>Appendix I – Curriculum Vitae</i></b>	<b>91</b>
<b><i>Appendix II – Publications</i></b>	<b>95</b>

## Introduction

### Celiac Disease

Celiac Disease (CD) is a chronic, small-intestinal enteropathy, which is triggered by gluten proteins contained in wheat, barley and rye. CD is characterized by an autoimmune response in genetically susceptible individuals resulting in small-intestinal mucosal injury. As a consequence, malabsorption develops, which in turn results in malnutrition-related problems including anemia, vitamin deficiencies, osteoporosis, and neurological disorders. A gluten-free diet (GFD) is sufficient to treat the overwhelming majority of patients with CD. The currently estimated prevalence is 1%, with a statistical range of probability of 0.5–1.26% in the general population in Europe and the USA (Dubé et al., 2005). Even taking into account that the actual occurrence rate of CD has been underestimated for many decades, the prevalence of this disease is increasing.

The spectrum of clinical manifestations is wide, most of them are secondary to malabsorption, as well as the natural history is variable. This prompted the Associazione Italiana Celiachia (AIC) to classify the disease, distinguishing four different clinical forms:

- *Typical*, early-onset (6-24 months) in a predominantly intestinal symptoms;
- *Atypical*, late-onset (6-7 years) and mostly extra-intestinal symptoms;
- *Silent*, which is CD without major digestive symptoms and so occasionally diagnosed in patients in apparent good health;
- *Potential*, characterized by a positive serological markers, but with normal histology.

Several environmental, genetic and socioeconomic factors contribute to the development of CD.

- *Infant feeding*: The role of infant feeding on the development of CD has been intensely debated since the late 1980s, which has resulted in a recommendation by the European Society for Pediatrics Gastroenterology, Hepatology and Nutrition (ESPGHAN) committee. This committee currently recommends that small amounts of gluten have to be gradually introduced between 4 and 7 months of age during breastfeeding (Agostoni et al., 2008).
- *Infections*: Infections after birth have been proposed to contribute to the development of CD. Whereas the role of infection with adenovirus type 12 in this process remains controversial, the association of Hepatitis C Virus infection and CD is well documented (Plot et al., 2009).
- *Socioeconomic features*: An epidemiological survey where comparisons were made between schoolchildren living in a prosperous area of Finland and children living in an adjacent poor region of Russia, whom in part shared genetic susceptibility and gluten intake, has suggested that worse socioeconomic conditions might protect against CD development (Kondrashova et al., 2008).

### ***Diagnosis***

At present the most sensitive and specific serological tests for diagnosis of CD are assessments of the presence of IgA auto-antibodies against the

endomysium of connective tissue (EMA) and against tissue transglutaminase (tTG) (Rostami et al., 1999).

In general, in case of strong clinical suspicion of CD, duodenal biopsy must be performed regardless of serological analysis; in cases of low suspicion of disease or screening, duodenal biopsy only needs to be performed in seropositive patients.

Although HLA-DQ2 and/or HLA-DQ8 positivity is not an absolute requirement for diagnosis, as 40% of the healthy western population also carry genotypes for these molecules, CD is highly unlikely in case both of them are absent. As a consequence of this high negative predictive value for developing CD, HLA genotyping was proposed as a contributing element to diagnosis, in particular in the absence of villous atrophy (Kaukinen et al, 2002).

### ***Genetic risk factors***

A strong genetic component is suggested by:

- Familial aggregation, as the prevalence of CD is 10 times higher in first degree familiars (~10%) than in the whole population (1%) (Bevan et al., 1999; Petronzelli et al., 1997);
- Disease concordance in monozygotic twins (> 80%) than dizygotic (~10%) (Greco et al., 2002).

### ***HLA genes***

HLA alone can explain about ~40% of the genetic predisposition to disease (Bevan et al., 1999). The genetic risk factors are the genotypes encoding the HLA class II molecules HLA-DQ2 (encoded by HLA-

DQA1\*0501 and HLA-DQB1\*02) and HLA-DQ8 (encoded by HLA-DQA1\*0301 and HLA-DQB1\*0302). About 90% of individuals with CD carry the DQ2 heterodimer encoded either in cis or in trans, and almost all of the remaining patients express DQ8 (Sollid et al., 1989) (Figure 1).

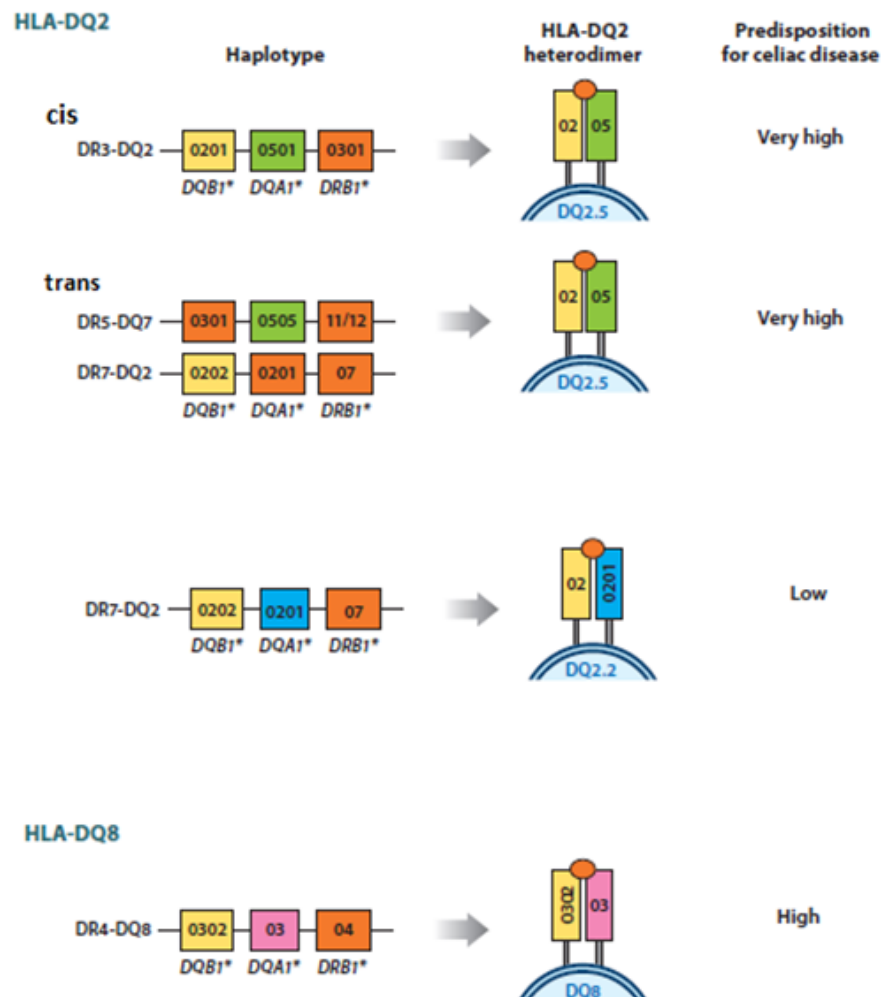


Figure 1. DQ2.5 haplotype is created by close genetic linkage of two alleles, written as a haplotype, DQA1\*0501:DQB1\*0201. The haplotype encodes DQ2.5cis isoform, referring to the *cis* arrangement of the DQA1\*0501 and DQB1\*0201 on the same variant of chromosome 6. The isoform can also be encoded trans-haplotype (between two sister chromosomes) forming the DQ2.5trans isoform. The DQ2.5 haplotype is linked to DR3. DQ2.2 is encoded almost exclusively by the DQA1\*0201:DQB1\*0202 haplotype. The haplotype is linked to DR7. DQA1\*0301:DQB1\*0302 (DQ8.1) is the most common DQ8 subtype representing over 98% of the DQ-8 bearing population.

Class II HLA molecules are localized on *Antigen Presenting Cells* (APC) surface and they are delegate to interact with T cell Receptor (TCR) in



order to present peptide molecules previously internalized by endocytosis, processed and brought to the surface. The molecules encoded by the genes of class II heterodimers are characterized by two transmembrane chains ( $\alpha$  and  $\beta$ ), and both chains are encoded by highly polymorphic genes (DQA and DQB, respectively) located in 6p chromosomal region. Most of the amino acid changes are located in a binding pocket which interacts with the antigenic peptide. The HLA polymorphism, therefore, affects the repertoire of peptides that can be presented to T lymphocytes and modulates immune response of each individual.

Human Tissue Transglutaminase (TGase) has a very important role in promoting intestinal damage: it is present at the basal membrane of intestinal epithelial cells and can deamidate gliadin peptides which have a higher binding affinity to HLA-DQ2 and HLA-DQ8 molecules than native ones. Deamidation produces glutamic acid, rich in negative charges, from the neutral original Glutamine; since the DQ2 molecule preferentially accommodates in its binding site peptides leading negative residuals in their structure, deamidated gliadin peptides can cause a strongest response by binding with higher affinity to the HLA-DQ2 molecule (van de Wal et al., 1998).

### *Non-HLA genes*

As it has been said previously, CD has a strong genetic component, more than others complex disease, and HLA genetic predisposition can explain only ~40% of the susceptibility to the disease. Currently, several susceptibility loci not related to HLA have been identified by genome-wide association studies (GWAs); these studies aimed to analyze two distinct populations of subjects: one (cases) characterized by presence of the disease, the other (controls) characterized by its absence. The hypothesis that supports an association study is that the presence of specific

polymorphisms in a locus (in particular Single Nucleotides Polymorphisms, SNPs) increases or decreases the risk to develop a complex diseases. There are in fact allelic variants that have a *predisposing role* to disease because they are more frequent in the population of affected individuals than in healthy controls, on the contrary there are variants with *protective role* that is less frequent in cases compared with healthy patients.

The first GWAs showed that the 4q27 region, which harbors the IL2, IL21 and KIAA1109 genes, had the strongest association with CD after the HLA region (van Heel et al., 2007). The 4q27 region was immediately investigated also in other auto-immune disease and an association was found also with type I Diabetes (Todd et al., 2007), Psoriasis (Liu et al., 2008) and Rheumatoid Arthritis (Zhernakova et al., 2007).

In 2008 the same group performed a new GWAs on a case-controls cohort confirming the association of the 4q27 region and finding other seven regions in which there are nine candidate SNPs (Hunt et al., 2008). According to them, all the new 8 loci could explain ~3-4% of the 60% unexplained by HLA.

From this point onwards, many replication on different population were performed in Europe, confirming or not for each SNP the association previously reported (Amundsen et al., 2010; Dema et al., 2009; Koskinen et al., 2009; J Romanos et al., 2009).

In 2009 others two SNPs were supposed to be associated to CD (Trynka et al., 2009); both are localized in or in proximity of two genes encoding two proteins (c-REL and TNFAIP3) involved in NF- $\kappa$ B pathway, important inflammation mediators.

With the last GWAs 13 risk variants were confirmed to be associated with CD, 13 new risk variants are proposed to be associated and further 13

met 'suggestive' criteria for association because of a little less significant p-value: 39 risk variants in total (Dubois et al., 2010; Table 1).

Chromosome	Position	SNP	Gene of interest
<b>Previously reported risk variants</b>			
1	190803436	rs2816316	RGS1
2	61040333	rs13003464	REL, AHSA2
2	102437000	rs917997	IL18RAP, IL18R1, IL1RL1, IL1RL2
2	181704290	rs13010713	ITGA4, UBE2E3
2	204510823	rs4675374	CTLA4, ICOS, CD28
3	46210205	rs13098911	CCR1, CCR2, CCRL2, CCR3, CCR5, CCR9
3	161147744	rs17810546	IL12A
3	189595248	rs1464510	LPP
4	123334952	rs13151961	IL2, IL21
6	32713862	rs2187668	HLA-DQA1, HLA-DQB1
6	138014761	rs2327832	TNFAIP3
6	159385965	rs1738074	TAGAP
12	110492139	rs653178	SH2B3
18	12799340	rs1893217	PTPN2
<b>New loci, genome-wide significant evidence</b>			
1	2516606	rs3748816	TNFRSF14, MMEL1
1	25176163	rs10903122	RUNX3
1	199158760	rs296547	?
2	68452459	rs17035378	PLEK
3	32990473	rs13314993	CCR4
3	120601486	rs11712165	CD80, KTELC1
6	90983333	rs10806425	BACH2, MAP3K7
6	128320491	rs802734	PTPRK, THEMIS
8	129333771	rs9792269	?
10	80728033	rs1250552	ZMIZ1
11	127886184	rs11221332	ETS1
16	11311394	rs12928822	CIITA, SOCS1, CLEC16A
21	44471849	rs4819388	ICOSLG
<b>New loci, suggestive evidence</b>			
1	7969259	rs12727642	PARK7, TNFRSF9
1	61564451	rs6691768	NFIA
1	165678008	rs864537	CD24
1	170977623	rs859637	FASLG, TNFSF18, TNFSF4
3	69335589	rs6806528	FRMD4B
3	170974795	rs10936599	?
6	328546	rs1033180	IRF4
7	37341035	rs6974491	ELMO1
13	49733716	rs2762051	?
14	68347957	rs4899260	ZFP36L1
17	42220599	rs2074404	?
22	20312892	rs2298428	UBE2L3, YDJC
X	12881445	rs5979785	TLR7, TLR8

Table 1: Genomic regions with the strongest association signals for celiac disease (Dubois et al., 2010)

The previous 13, the new 13 and the 13 new suggestive risk variants define the whole inflammation and immunological pathways, as shown by

the DAVID results for Biological Process (<http://david.abcc.ncifcrf.gov/>; Table 2). Candidate genes were grouped for Gene Ontology (GO) category and False Discovery Rate (FDR) has been fixed at  $<0.05$ . In a list of statistically significant findings (i.e. studies where the null-hypothesis could be rejected), FDR procedures are designed to control the expected proportion of incorrectly rejected null hypotheses ("false discoveries") (Benjamini & Hochberg, 1995).

Term	Count	P-value	Genes	FDR
GO:0006955~immune response	24 (43.6%)	3.64E-17	HLA-DQB1, CIITA, IL18RAP, TNFSF4, THEMIS, CCR1, CTLA4, FASLG, TNFRSF14, TLR7, TNFSF18, TLR8, HLA-DQA1, CCR9, RGS1, CCR5, CCR4, ETS1, ICOS, CCR2, IL12A, ICOSLG, CD28, IL2	5.63E-14
GO:0050863~regulation of T cell activation	10 (18.2%)	3.69E-10	MAP3K7, CD80, IL12A, CTLA4, TNFRSF14, IRF4, IL21, ICOSLG, CD28, IL2	5.71E-07
GO:0050865~regulation of cell activation	11 (20.0%)	6.66E-10	MAP3K7, PLEK, CD80, IL12A, CTLA4, TNFRSF14, IRF4, IL21, ICOSLG, CD28, IL2	1.03E-06
GO:0042129~regulation of T cell proliferation	8 (14.5%)	2.57E-09	CD80, IL12A, CTLA4, TNFRSF14, IL21, ICOSLG, CD28, IL2	3.97E-06
GO:0051249~regulation of lymphocyte activation	10 (18.2%)	3.01E-09	MAP3K7, CD80, IL12A, CTLA4, TNFRSF14, IRF4, IL21, ICOSLG, CD28, IL2	4.65E-06
GO:0002694~regulation of leukocyte activation	10 (18.2%)	8.26E-09	MAP3K7, CD80, IL12A, CTLA4, TNFRSF14, IRF4, IL21, ICOSLG, CD28, IL2	1.28E-05
GO:0002684~positive regulation of immune system process	11 (20.0%)	1.30E-08	MAP3K7, CD80, THEMIS, CD247, IL12A, IL21, TLR7, TLR8, ICOSLG, CD28, IL2	2.02E-05
GO:0001817~regulation of cytokine production	10 (18.2%)	1.76E-08	MAP3K7, TNFSF4, CD80, REL, IL12A, IRF4, IL21, TLR7, TLR8, CD28	2.72E-05
GO:0042108~positive regulation of cytokine biosynthetic process	7 (12.7%)	2.04E-08	CD80, REL, IRF4, IL21, TLR7, TLR8, CD28	3.15E-05
GO:0050670~regulation of lymphocyte proliferation	8 (14.5%)	2.04E-08	CD80, IL12A, CTLA4, TNFRSF14, IL21, ICOSLG, CD28, IL2	3.16E-05
GO:0070663~regulation of leukocyte proliferation	8 (14.5%)	2.22E-08	CD80, IL12A, CTLA4, TNFRSF14, IL21, ICOSLG, CD28, IL2	3.44E-05
GO:0032944~regulation of mononuclear cell proliferation	8 (14.5%)	2.22E-08	CD80, IL12A, CTLA4, TNFRSF14, IL21, ICOSLG, CD28, IL2	3.44E-05
GO:0050867~positive regulation of cell activation	8 (14.5%)	1.55E-07	MAP3K7, PLEK, CD80, IL12A, IL21, ICOSLG, CD28, IL2	2.39E-04
GO:0042035~regulation of cytokine biosynthetic process	7 (12.7%)	2.86E-07	CD80, REL, IRF4, IL21, TLR7, TLR8, CD28	4.43E-04
GO:0042102~positive regulation of T cell proliferation	6 (10.9%)	2.94E-07	CD80, IL12A, IL21, ICOSLG, CD28, IL2	4.55E-04
GO:0050870~positive regulation of T cell activation	7 (12.7%)	3.36E-07	MAP3K7, CD80, IL12A, IL21, ICOSLG, CD28, IL2	5.19E-04

Table 2: DAVID results of associated genes located in 39 risk loci: Gene Ontology for Biological Processes. GO (Gene Ontology) is the number concatenated to category name (es. GO:0006955 is the category related to the immune response); (FDR) False discovery rate control is a statistical method used in multiple hypothesis testing to correct for multiple comparisons.

To simplify, the candidate genes were included in four categories:

1. *T-cell development in the thymus*: the associated variants might alter biological processes before thymic MHC-ligand interactions.
  - *THEMIS* has a key regulatory role in both positive and negative T-cell selection during late thymocyte development.
  - *RUNX3*, a master regulator of CD8<sup>+</sup> T lymphocyte development in the thymus.
  - *TNFRSF14* has widespread functions in peripheral leukocytes and a crucial role in promoting thymocyte apoptosis.
  - The *ETS1* transcription factor is also active in peripheral leukocytes; however, it is also a key player in thymic CD8<sup>+</sup> lineage differentiation, acting in part by promoting RUNX3 expression.
2. *Innate immune detection of viral RNA*.
  - *TLR7-TLR8* recognize viral RNA and viral infection (and the nature of the host response to infection) as a putative environmental trigger that could be common to these autoimmune diseases.
3. *T- and B-cell co-stimulation (or co-inhibition)*.
  - The association of genes like CTLA4- ICOS-CD28, TNFRSF14, CD80, ICOSLG, TNFRSF9, TNFSF4 indicate that fine control of the adaptive immune response might be altered in individuals at risk of CD.
4. *Cytokines, chemokines and their receptors*.

These confirmed susceptibility loci can explain approximately 20% of disease variance (Dubois et al., 2010) and help refine the pathways

involved in CD pathogenesis. Recently a new method using polygenic risk-score analyses to infer the total liability-scale variance explained by associated GWAS SNPs was used to refine this value, and for CD the value obtained was 43% (excluding MHC contribute) (Stahl et al., 2012). Nonetheless, much more work needs to be done to explain the missing heritability and to understand the functional consequences of these risk alleles.

As it's almost impossible to study all of them in a thorough way, we decided to focalize our attention on a few, in particular those which resulted to be associated with CD in different association studies on different populations and so less amenable to an environmental effect.

#### *Function of selected candidate genes*

Firstly we pointed attention on the 4q27 locus, because of its redundant positive result in all GWAs performed until now. Three main genes belong to this locus: KIAA1109, IL-2 and IL-21.

Recent studies showed that the encoded protein of **KIAA1109** gene (RefSeq Accession NP\_056127.2) is similar to a Chinese hamster protein associated with spermatocyte and adipocyte differentiation (Wei et al., 2006). The C-terminus of the protein is also similar to a *Caenorhabditis elegans* protein that plays a role in lipid storage (Wei et al., 2006). In mammals, this protein is thought to have a role in the regulation of epithelial growth and differentiation, and in tumor development (Kuo et al., 2006). The literature about KIAA1109 is very scarce, so it's not well known how it could correlate with CD, even if this gene belongs to the 4q27 locus which is the first and most associated one, after HLA, not only to CD (van Heel et al., 2007) but also to type I diabetes (Todd et al., 2007),

Psoriasis (Liu et al., 2008) and Rheumatoid Arthritis (Zhernakova et al., 2007).

**IL-2** is produced from T cell after antigenic or mitotic stimulation and it works in autocrine or paracrine manner on monocytes and on T, B, LAK, NK and glioma cells. The decreased expression of IL-2 is associated to an organ-specific autoimmune reaction, because of its capacity to activate Treg cell (CD4<sup>+</sup> CD25<sup>+</sup>) and so inhibiting the negative selection of T cell clones reactive towards self (Yamanouchi et al., 2007).

**IL-21** is one of the crucial cytokine in the inflammatory and gluten induced immune response. Its pro-inflammatory action directly correlates with the increase of IFN- $\gamma$  (Fina et al., 2008). Moreover, together with IL-15 promotes NK cell proliferation and maturation, and itself alone has the same effect on T and B cell as an answer to activating stimuli (De Nitto et al., 2009).

**c-REL** is a subunit of the NF- $\kappa$ B complex, and it is crucial for the regulation of this major nuclear factor which regulates the innate and adaptive immunity (Bonizzi et al., 2004). It's regulation can terminate the NF- $\kappa$ B activation by innate immune stimuli. The integral structure of c-REL is crucial for the activation as well as for the termination of NF- $\kappa$ B induced immunoresponse (Lawrence et al., 2005). Our group produced evidences that NF- $\kappa$ B is also constitutively active in intestinal mucosa of patients with untreated CD and reverts to normal values when gluten is removed from the diet (Maiuri et al., 2003).

**LPP** (Lipoma Preferred Partner) gene encodes for a protein not directly implicated in immune and/or inflammatory disorders but according to recent GWAs the SNP association value is one of the highest, and so it seems to be more closely involved in the pathogenesis of CD than the others (Dubois et al., 2010); moreover this is the most associated gene also



in a family study, as reported by us (Izzo et al., 2011). It is a member of LIM domain-proteins family (Petit et al., 1996) probably involved in the regulation of shape, in cell polarity and motility (Petit et al., 2000), and in zinc finger protein transfer through cell-cell adhesions (Petit et al., 2003).

**RGS1** (Regulator of G-protein signaling 1) gene encodes a member of the “G protein signaling regulator” family which is probably involved in homing mechanism, because *Rgs1*<sup>-/-</sup> mice showed an increased B cell movement from and towards lymph nodes (Han et al., 2005) and an increased migratory response of dendritic cells after chemokines stimulation (Shi et al., 2004).

About **TNFSF14** gene, it is located on 19p13.3 locus, and the protein encoded is a member of tumor necrosis factor (TNF) ligands superfamily; its receptor (**TNFRSF14**), also known as Herpesvirus Entry Mediator (HVEM), resulted to be associated with CD in one of the latest GWAS (Dubois et al., 2010). In murine models, the gene is constitutively expressed in inflamed intestinal mucosa (Shaikh et al., 2001). Moreover, recent experiments have shown that there is a high expression in human intestinal T cells, and that its expression is essential for the production of IFN- $\gamma$  (Cohavy et al., 2004).

**SH2B3** gene encodes for a protein capable to negatively regulate the lymphopoiesis and early hematopoiesis, and its down-regulation can lead to increased production of B cells, and increased expansion and enhanced function of hematopoietic stem cells (HSCs) (Takaki, 2008); it is a negative regulator of PDGFR signaling (Gueller et al., 2011).

**TAGAP** (*T-cell activation GTPase activating protein*) gene encodes a protein expressed in activated T cells and it's predicted to function as a Rho GTPase-activating protein (Chen et al., 2011).

With regard to **TNFAIP3** gene, Tnfaip3<sup>-/-</sup> mice had greater intestinal permeability compared to wild-type (Kolodziej et al., 2011), suggesting that this protein can maintain intestinal barrier function and support epithelial cell tight junctions. It is also closely linked to NF-κB pathway because failure to down regulate NF-κB transcriptional activity resulted in chronic inflammation and cell death, as observed in A20-deficient mice making TNFAIP3 a potent inhibitor of NF-κB signaling (Wertz et al., 2004).

## Aim

The aim of this study is to improve the knowledge about the role of the genetics in the CD context, in order to see non-HLA genes not like a list of genes with certified immunological function, but to understand the reasons why they are associated with CD.

For this reason the project of the thesis is articulated in several sub-projects.

1. *Gene Expression studies and non-HLA genes*: our aim to improve the analyses of candidate gene expression comparing the results obtained on mucosal biopsies, peripheral monocytes and T-cell lines isolated from duodenal mucosa: three different cell populations with three different functions and locations, in order to give a picture of the variation in the celiac type and to correlate it with a model of CD pathogenesis.

Moreover, a discriminant analysis of the celiac gene expression profile has recently been proposed as a promising diagnostic tool with which to distinguish celiac mucosa from normal mucosa (Bragde et al.,2011). Therefore, we evaluated the diagnostic strength of the discriminant analysis on the gene expression data obtained from peripheral blood monocytes with the aim to provide a less invasive alternative to duodenal biopsy as the only diagnostic tool for CD.

2. *Non-HLA genes and Potential CD patients*. Our aim was to explore the presence of genetic and expression factors that may differentiate potential patients from overt CD patients and controls, in particular:

- To find a different distribution of HLA haplotypes;

- To evaluate if there is a different distribution of non-HLA genes;
- To confirm association results also in expression studies on RNA extracted from intestinal biopsies.
- To understand if a correlation exists between genotype and expression levels.

3. *Non-HLA genes as a model for risk prediction in a family cohort.*

The aim of this study was to study a genetic risk model for CD using HLA and non-HLA risk alleles. We aimed to significantly improve the identification of subjects at risk to develop CD by modeling the HLA risk with the newly discovered genetic profile in order to reach an higher accuracy in prediction. When this goal will be reached, families with one case will know, with reliable precision, the individual risk of a newborn and will be able to anticipate the onset of the clinical phenotype by a simple point-of-care tests. After families the model might well be extended to populations, in the frame of national policies to screen for genetic predispositions to diseases.

## **Chapter 1**

**Non-HLA genes and Celiac Disease:**

**Gene expression studies**

## Chapter 1. Non-HLA genes and Celiac Disease: Gene expression studies

### Background and aim

GWA studies represent a major breakthrough in the genetics of complex disorders. At this point, besides the HLA contribution, we know that 39 genes are associated to CD, but no one knows how these polymorphic variants can contribute to the pathogenesis of the disease. Dubois and his group (Dubois et al., 2010) performed the largest GWA study in CD history; they analyzed 1.469 whole blood samples in an expression quantitative trait meta-analysis: 20 out of 38 (52.6%) tested loci had celiac risk variants correlated with *cis* gene expression (Figure 2).

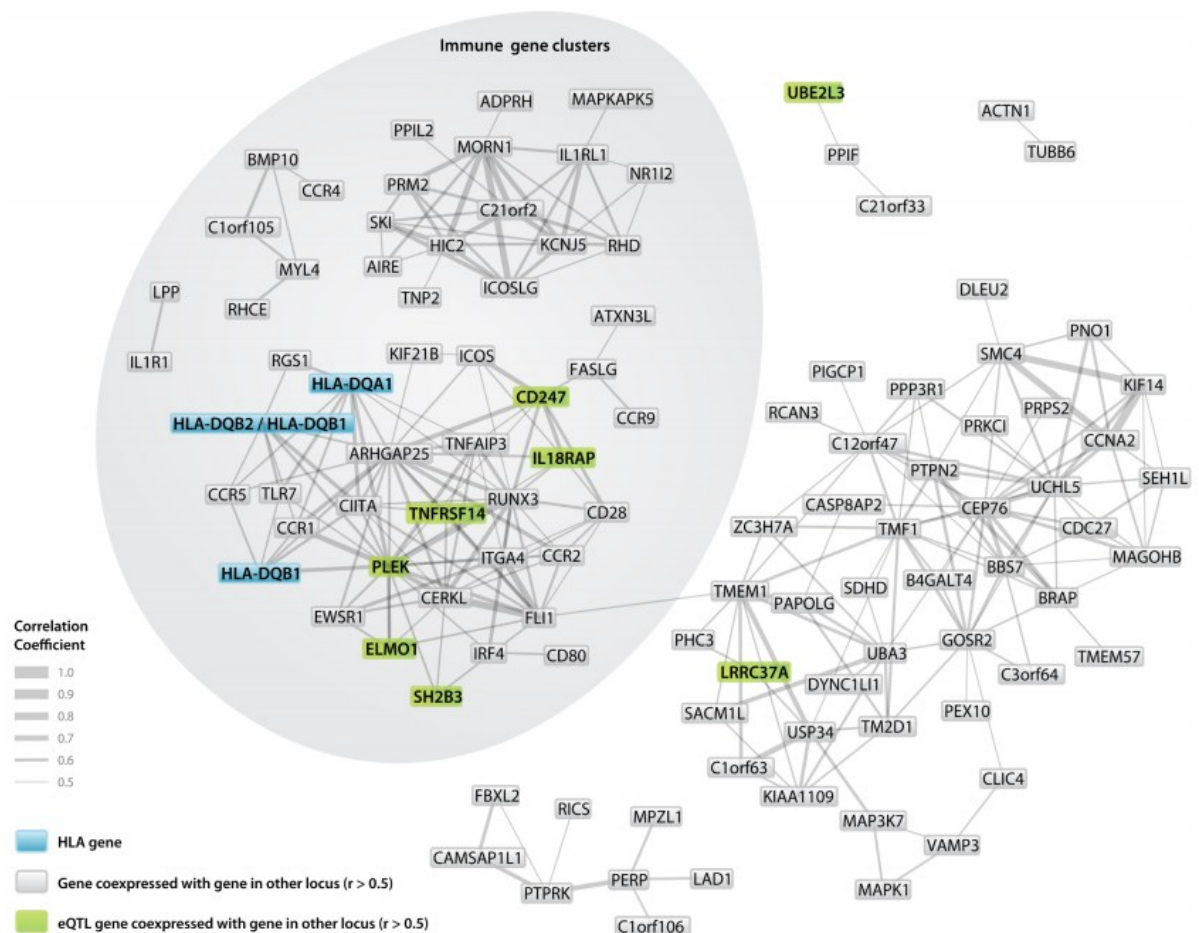


Figure 2: Co-expression analysis of genes mapping in 39 genome-wide significant and suggestive CD associated regions in 33,109 heterogeneous human samples from the Gene Expression Omnibus (Dubois et al., 2010).

The problem is that these analyses were performed on whole blood, but literature about an effective correspondence between the association data and the gene expression profile in the target tissue is still missing, despite it could be useful to explore the mechanisms behind the atrophic duodenal outcome. In a recent study (Plaza-Izurieta et al., 2011) in which the expression analysis of associated genes was performed on duodenal mucosa samples in a Spanish celiac population, correlating it with the tagSNP genotype, although the study does not always confirm the previous results of eQTL profile (Dubois et al., 2010).

Moving from these assumptions, our aim is to improve the analyses of candidate gene expression comparing the results on mucosal tissue with those obtained on peripheral monocytes and T-cell lines isolated from duodenal mucosa: three different cell populations with three different functions and locations, in order to give a picture of the variation in the celiac type and to correlate it with a model of CD pathogenesis.

Moreover, a discriminant analysis of celiac gene expression profile has recently been proposed as a promising diagnostic tool to distinguish celiac mucosa from normal mucosa (Bragde et al., 2011). However, only genes involved in the alteration of crypt-villi architecture were evaluated and only duodenal tissue was tested. Therefore, we evaluated the diagnostic strength of the discriminant analysis on the gene expression data obtained from peripheral blood monocytes with the aim of providing a less invasive alternative to duodenal biopsy as the only diagnostic tool for CD.

Genes were selected from those resulted to be associated with CD in different studies on different populations and so less amenable to an environmental effect, in particular KIAA1109, IL-2, IL-21, c-REL, LPP, RGS1, SH2B3, TAGAP, TNFAIP3 and TNFRSF14. An additional evaluation was also

performed on TNFSF14 mRNA because of the CD-association of its receptor TNFRSF14.

## **Materials and methods**

### ***Patients***

For gene expression analysis, duodenal biopsies were obtained during gastroduodenal endoscopy (GDE) procedures and fresh-frozen in liquid nitrogen. At least 10 patients with untreated CD, 3 patients with treated CD and 5 non-celiac patients were analyzed. The diagnosis of CD was based on ESPGHAN criteria (Working Group of European Society of Paediatric Gastroenterology and Nutrition, 1990). Controls included patients with a normal duodenal with no mucosal atrophy (Marsh stage of lesion M0, (Oberhuber et al., 2001)), they underwent to GDE because of Gastritis, Gastroesophageal reflux disease or hypothesis not confirmed of *Helicobacter Pylori* infection.

### ***Monocytes***

We used the Dynabeads® My Pure™ Monocyte kit (Life Technologies, Foster City, CA) to isolate monocytes from other peripheral blood cell types (B- and T-lymphocytes, NK cells, erythrocytes, dendritic cells etc.). Monocytes were extracted from peripheral blood of 18 control patients, 8 CD patients, 11 Crohn disease patients and 5 celiac patients on a GFD. To verify the discriminant results, we analyzed a second cohort of 9 CD patients.



### ***Gliadin preparation***

Gliadin was extracted from a common exaploid wheat cultivar (Sagittario) according to a standard procedure and enzymatically digested with pepsin and trypsin, as previously described (Gianfrani et al., 2007). PT-gliadin was deamidated with guinea pig tTG.

### ***Generation of gliadin-specific T-cell lines and functional assay***

Gliadin-reactive TCLs were generated from duodenal explants of 3 adult control, (non celiacs) and 3 untreated HLA-DQ2<sup>+</sup> celiacs, as described previously (Gianfrani et al., 2007). Briefly, mucosal explants were digested with collagenase-A and cells were suspended at  $2-3 \times 10^5$ /mL in complete medium (X-Vivo15 medium supplemented with 5% AB pooled human serum and antibiotics, all provided from BioWhittaker, Verviers, Belgium). Mucosal cells were stimulated with  $1.5 \times 10^6$  irradiated autologous PBMCs, deamidated-PT-gliadin (50 µg/mL), and IL-15 and IL-2 as growth factors (R&D System, Minneapolis, MN). After 7 and 21 TCLs were re-stimulated with deamidated-PT-gliadin and feeder cells. For gene expression analysis, TCLs ( $0.5-1.0 \times 10^6$ ) were stimulated with autologous, EBV-transformed B lymphoblastoid cell lines (B-LCLs,  $1.0 \times 10^6$ ) pulsed overnight in presence or absence of deamidated-PT-gliadin (100 µg/mL). After 0, 1, 3 hours of incubation, cells were pelleted and cryopreserved at -80°C. Before cryopreserving, an aliquot of cell supernatants were harvested for IFN-γ measurement.

### ***Expression studies***

From frozen tissues samples, total RNA was isolated using RiboPure™ kit (Life Technologies, Foster City, CA). The quality of the RNA was assessed by running aliquots on agarose gels. In each Real Time PCR experiment, double stranded complementary DNA (cDNA) was synthesized from 2 µg of total RNA using the High-Capacity cDNA reverse transcription kit (Life Technologies, Foster City, CA) as per the manufacturer's protocol. Retro-transcription had to be followed by a linear pre-amplification step before the Real Time experiment because of the poor amount of RNA recovered from monocyte samples. This step was performed using TaqMan® PreAmp Master Mix (Life Technologies, Foster City, CA) that preamplifies small amounts of cDNA without introducing amplification bias to the sample.

Experiments were performed on 7900HT Fast Real Time PCR using in 20µL of reaction TaqMan® Gene Expression Assay 20x (Life Technologies, Foster City, CA), TaqMan® Gene Expression Master Mix 2x (Life Technologies, Foster City, CA) and about 40ng of cDNA, as per the manufacturer's protocol; each gene expression was normalized towards an endogenous housekeeping (GUSb). Relative expression ratios were calculated as  $R = 2^{[Ct(GUSb) - Ct(test)]}$ , where R is ratio,  $C_t$  is the cycle number at the threshold, and test refers to the tested mRNA. The confidence interval was fixed at 95%. Data were normalized to the control group.

All primers and probes for each gene were purchased as commercial 'Assays-On-Demand' sets from Life Technologies (Foster City, CA), chosen because of their gene target position: probe binds and amplifies an exonic junction, which can permit us to exclude from our analysis any DNA genomic contaminants.

## ***Statistical Analysis***

The non-parametric Mann-Whitney U test for 2 independent variables was used to assess the difference between data sets. Statistical analyses were performed using SPSS 17.0 (SPSS Inc., Chicago, IL, USA) and GraphPad Prism 5.0 (GraphPad software, San Diego, CA, USA).

## **Results**

### ***Expression studies***

#### ***Mucosal tissue***

KIAA1109 and c-REL mRNA expression in intestinal mucosa of CD patients was not significantly different from controls, while the expression increased significantly in CD-GFD mucosa compared to controls ( $p=0.01$ , Mann-Whitney U: 13.00 and  $p<0.01$ , Mann-Whitney U: 5.000 respectively) and CD patients ( $p=0.03$ , Mann-Whitney U: 9.00 and  $p=0.01$ , Mann-Whitney U: 4.000 respectively) (Figure 3a-b). On the contrary, IL 21 gene was more expressed in CD patients than in controls ( $p=0.01$ , Mann-Whitney U: 3.00; Figure 3c) according to data reported in literature (Fina et al., 2008); this increase seems to be gluten-dependent because the expression returned to the levels of controls after at least one year of GFD (CD vs. CD-GFD  $p=0.01$ , Mann-Whitney U: 0.00). Also SH2B3 gene showed a mild not significant increase of gene expression in CD duodenal biopsies while in CD-GFD the expression decreased sharply in a significant way compared to both controls ( $p<0.01$ , Mann-Whitney U: 3.00) and CD ( $p<0.01$ , Mann-Whitney U: 2.00) (Figure 3d). In our studies TAGAP mRNA significantly increased in CD mucosal biopsies compared to controls

( $p=0.04$ , Mann-Whitney U: 23.00) (Figure 3e). Mucosal biopsies of CD patients showed higher values of TNFSF14 mRNA than controls ( $p<0.01$ , Mann-Whitney U: 8.00), and it is noteworthy that these levels remain high in CD-GFD patients when compared to Controls ( $p=0.04$ , Mann-Whitney U: 9.00) (Figure 3f).

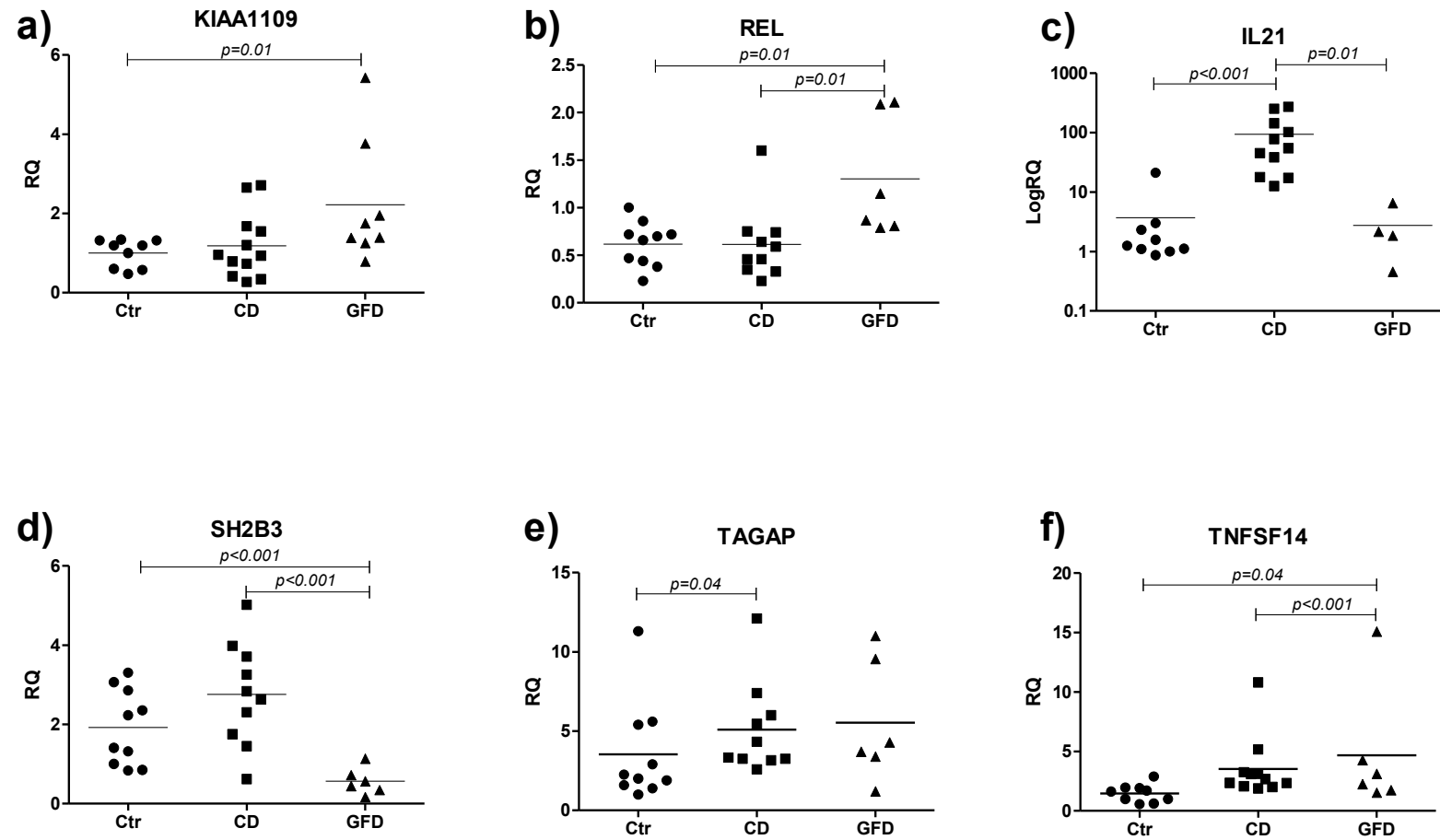


Figure 3. Gene expression studies on small intestinal biopsies.

The analysis of IL-2 mRNA showed that there were no significant differences among CD patients, controls and CD-GFD; in particular a trend of slight over-expression was shown in CD compared to Controls and CD-GFD patients. The analyses of LPP, RGS1, TNFAIP3 and TNFRSF14 showed no significant differences between controls, CD and CD-GFD patients, nevertheless RGS1 gene expression was slightly improved in CD compared to controls, even if this difference did not reach the significant threshold. All these last results are resumed in Figure 4.

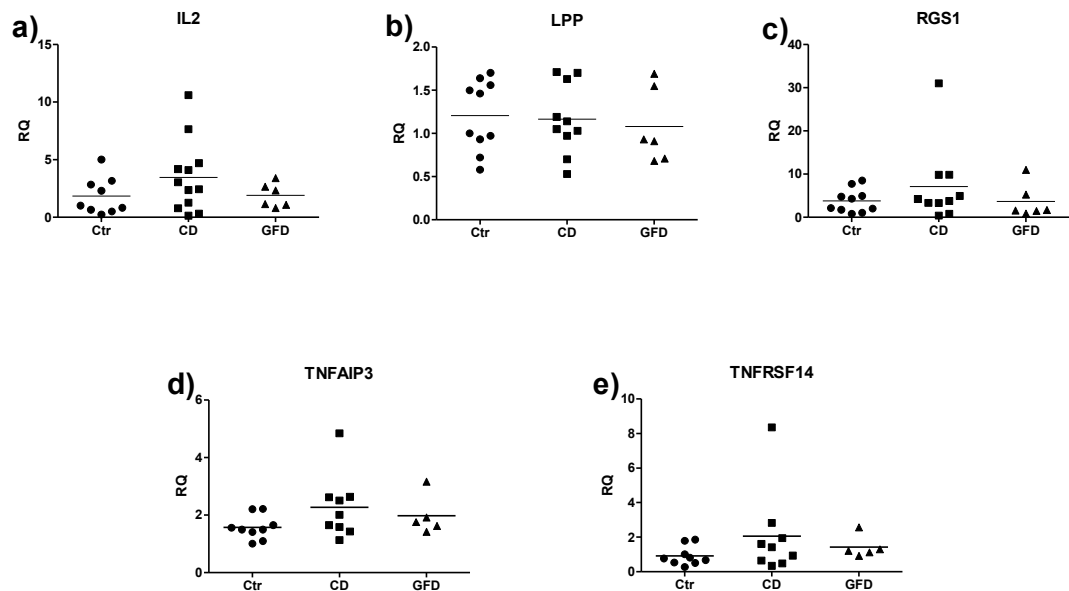


Figure 4. Gene expression studies on mucosal tissue of IL-2, LPP, RGS1, TNFAIP3 and TNFRSF14 genes.

#### *Peripheral blood monocytes*

The expression of the panel of candidate genes was evaluated in monocytes extracted from peripheral blood samples of 18 controls, 17 CD and 5 CD-GFD patients. The aim was to evaluate the differences with the results obtained in duodenal mucosa, and to investigate the involvement of these genes in peripheral processes also, for example, before the

recruitment of monocytes from the blood to intestinal tissue. Monocytes extracted from peripheral blood of 11 Crohn patients served as positive controls. We did not evaluate the expression of IL-2 and IL-21 because they are not produced by monocytes but by CD4<sup>+</sup> TCLs after antigen activation (De Nitto et al., 2009; Williams et al., 1991).

The KIAA1109 gene was over-expressed in CD (t-test  $p=0.05$ ; Means  $\pm$  SEM "0.8650  $\pm$  0.04832 N=18" vs. "1.068  $\pm$  0.08723 N=17"), Crohn (t-test  $p=0.02$ ; Means  $\pm$  SEM "0.8650  $\pm$  0.04832 N=18" vs. "1.435  $\pm$  0.2794 N=11") and CD-GFD patients versus controls (t-test  $p=0.05$ ; Means  $\pm$  SEM "0.8650  $\pm$  0.04832 N=18" vs. "1.110  $\pm$  0.1445 N=5"; Figure 4a). Differently, c-REL, SH2B3 and LPP expression was lower in CD monocytes compared with controls (c-REL –  $p<0.01$ ; Mann-Whitney U 11.50; SH2B3 – t-test  $p=0.05$ ; Means  $\pm$  SEM "1.097  $\pm$  0.1163 N=17" vs. "0.7500  $\pm$  0.1153 N=15"; LPP –  $p=0.05$ ; Mann-Whitney U 80.0), but regarding c-REL and SH2B3 genes significantly higher than in either CD-GFD (c-REL – t-test  $p<0.01$ ; Means  $\pm$  SEM "0.9612  $\pm$  0.08464 N=17" vs. "2.812  $\pm$  0.4869 N=5"; SH2B3 – t-test  $p<0.01$ ; Means  $\pm$  SEM "1.097  $\pm$  0.1163 N=17" vs. "2.448  $\pm$  0.2510 N=4") or Crohn monocytes (c-REL –  $p<0.01$ ; Mann-Whitney U: 10.0; SH2B3 – t-test  $p=0.01$ ; Means  $\pm$  SEM "1.097  $\pm$  0.1163 N=17" vs. "1.895  $\pm$  0.3358 N=11") (Figure 4b, 4c and 4e). The expression of RGS1 showed a trend similar to that of c-REL, but did not differ between CD-GFD and control monocytes (Figure 4d). The ANOVA revealed a significant difference between the groups in terms of mean RQ as regards c-REL, SH2B3 and TNFRSF14 expression. In all cases the p-value was lower than 0.01)

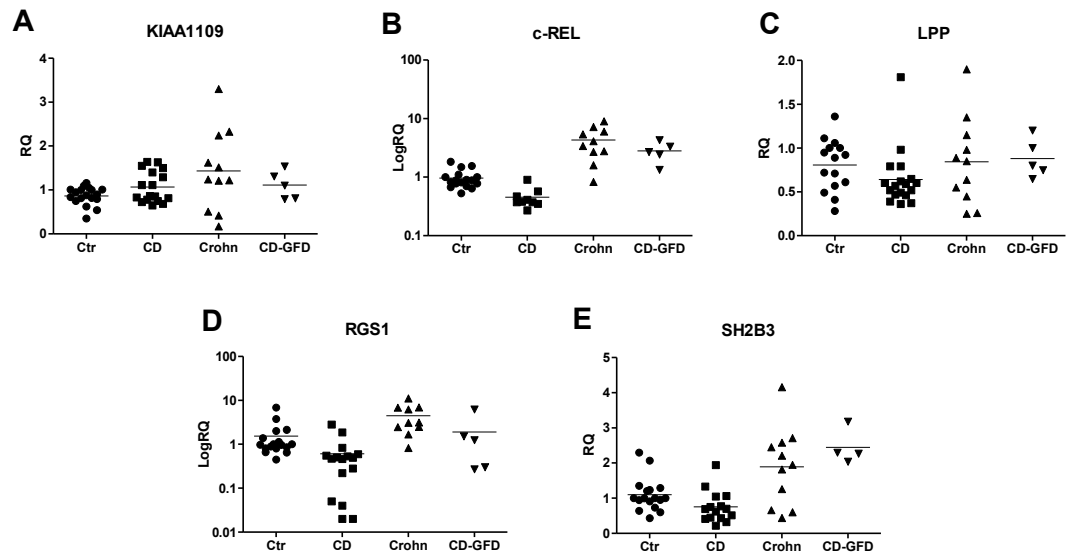


Figure 5. Gene expression studies on peripheral blood monocytes of KIAA1109, c-REL, LPP, RGS1 and SH2B3 genes.

TAGAP, TNFSF14 and TNFRSF14 were expressed at similar levels in CD and CD-GFD monocytes and results are resumed in Figure 6.

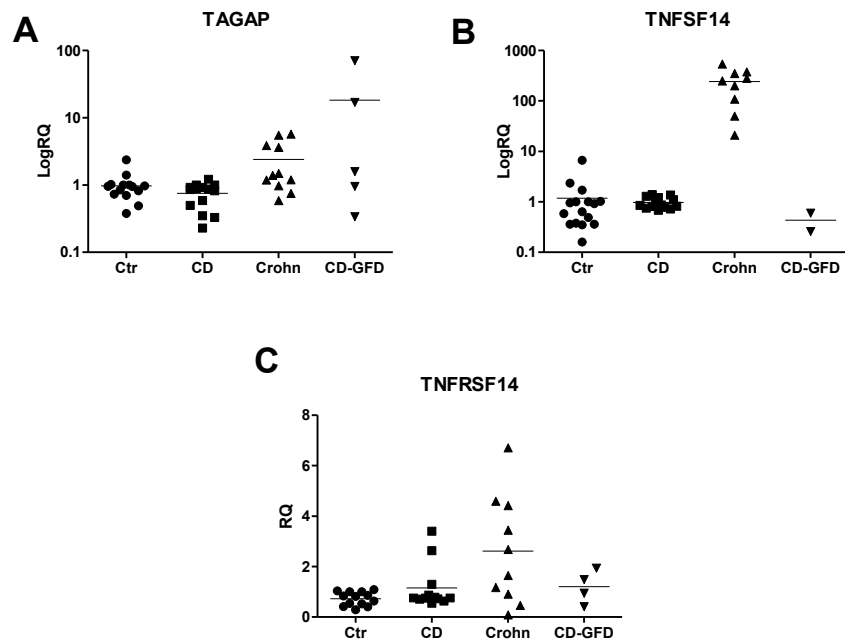


Figure 6. Gene expression studies on peripheral blood monocytes of TAGAP, TNFSF14 and TNFRSF14 genes.



*T-cell lines isolated from duodenal biopsies.*

We isolated CD4<sup>+</sup> TCLs from the duodenal mucosa of 3 controls and 3 CD patients, raised against deamidated PT-gliadin digest. Only CD-derived TCLs highly reactive to deamidated PT-gliadin were selected. To evaluate gene expression in terms of gliadin dependence and/or time dependence, TCLs were grown in a culture medium with or without deamidated PT gliadin and fixed at baseline (time 0) and after 1h and 3h of incubation. We labeled these conditions as follows: M0=medium culture at 0h; M1=medium culture after 1h; M3=medium culture after 3h; G0=gliadin culture at 0h; G1=gliadin culture after 1h-incubation; G3=gliadin culture after 3h-incubation.

Samples were analyzed at time 0h and 3h to determine at which time-point the expression changes occurred. The expression at 1 h was evaluated only if differences were found between baseline and 3 h. As expected, there was no difference in terms of gene expression in controls at time 0, incubated with or without gliadin (controls M0 vs controls G0), or in CD samples at time 0, incubated with or without gliadin (CD M0 vs CD G0).

IL-2 expression differed significantly between CD and controls already at time 0, irrespective of the presence of gliadin in the culture medium (CD M0/G0 vs controls M0/G0). After 1 and 3h of incubation, gliadin significantly stimulated IL-2 mRNA expression, whereas at the same time points, in the absence of gliadin, IL-2 expression was significantly lower in CD than in control samples (Figure 7A). IL-21 expression at baseline was similar in CD and controls in both T-cell populations incubated with and without gliadin (CD M0/G0 vs controls M0/G0). After exposure to gliadin for 1h and for 3h, IL-21 expression was higher in the CD group than in controls (CD G1 vs controls M0/G0 and CD M0/G0). Interestingly, IL-21

expression was higher in celiac TCLs after incubation for 3h in medium alone (CD M3 vs controls M0; Figure 7B). Similar to IL-21, the expression of c-REL at baseline in CD and in controls did not differ between the two T-cell populations incubated with and without gliadin (CD M0/G0 vs controls M0/G0). At 1h, c-REL expression increased slightly in control and CD TCLs, irrespective of gliadin (CD M1/G1vs controls M0/G0), whereas after 3h of incubation, c-REL expression was higher in CD than in control TCLs, and the difference was more pronounced in the presence of gliadin (control M3/G3 vs CD M3 and G3) (Figure 7C). TNFAIP3 expression was similar to that of c-REL, except for an increase in expression in CD TCLs after 1h of incubation (Figure 7D). RGS1 gene expression was much lower in CD than in control TCLs at time 0, irrespective of the presence of gliadin in the medium (CD M0/G0 vs controls M0/G0); it remained lower also after 1h and 3h of incubation with medium alone. After 3h of incubation with gliadin, the expression of TNFAIP3 was up-regulated and reached a level comparable to that of controls in the same condition (Figure 7E).

No difference in the expression of any gene was found in control TCLs up to three hours of incubation with and without gliadin (data not shown).

All the significant differences have positive Mann-Whitney test ( $p=0.05$ ).

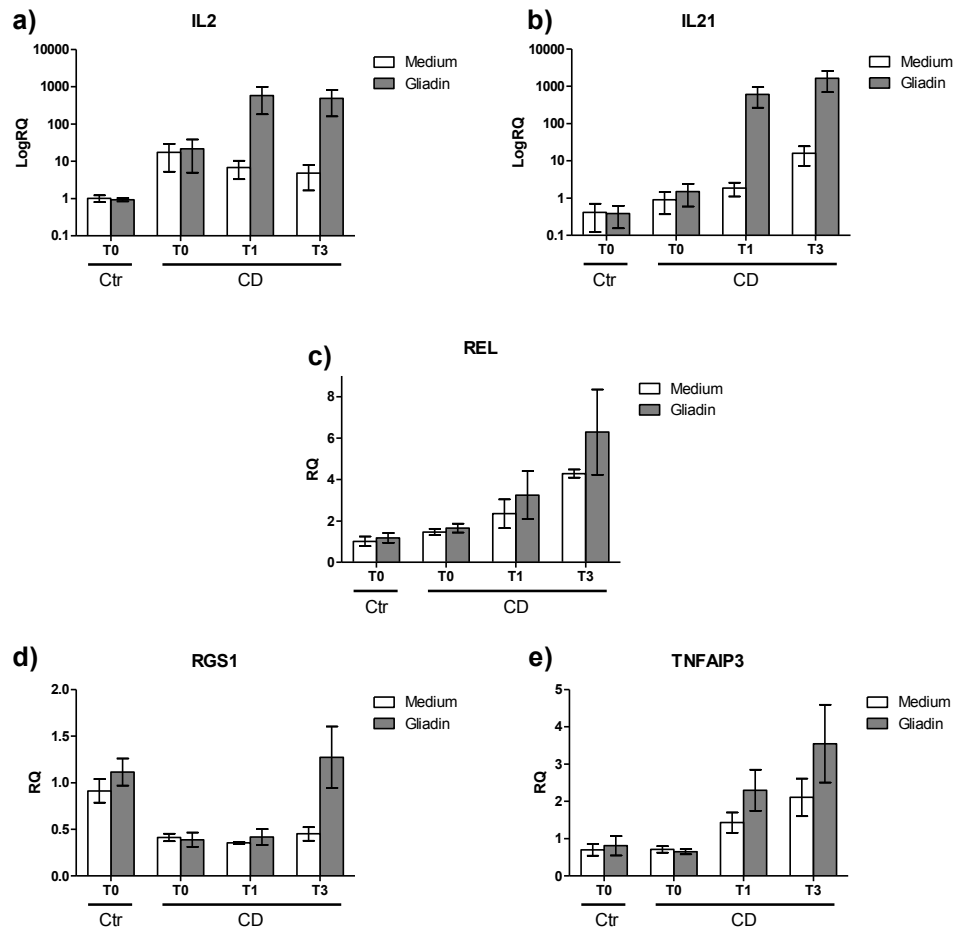


Figure 7: gene expression studies on T-cell lines isolated from duodenal biopsies of IL-2, IL-21, c-REL, RGS1 and TNFAIP3 genes.

The expression of KIAA1109, LPP, SH2B3, TAGAP, TNFSF14 and TNFRSF14 did not differ between CD in samples incubated with and without gliadin measured at 0h, and 3h. Hence we did not measure their expression after 1h of incubation (Figure 8).

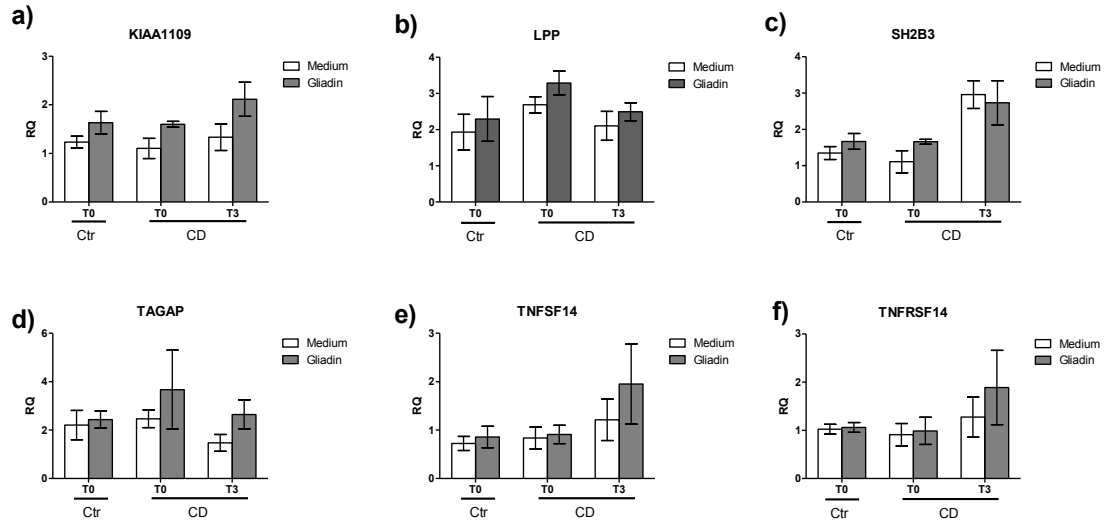


Figure 8: gene expression studies on T-cell lines isolated from duodenal biopsies of KIAA1109, LPP, SH2B3, TAGAP, TNFSF14 and TNFRSF14 genes.

## Statistical Analyses

### Discriminant Analysis in small intestinal tissue

In order to identify genes whose expression in small intestinal mucosa better characterize celiac tissue versus controls, a linear discriminant equation was fitted to the standardized values of expression (RQ). At each step, the variable that minimizes the overall Wilks' Lambda is entered. The following limits were used for the analysis:

- Maximum number of steps is 12
- Minimum partial F to enter is 1.84
- Maximum partial F to remove is 1.71

According to this analysis, 5 genes (stepwise: TNFAIP3, IL-21, c-REL, RGS1 and LPP) were selected from those analyzed, with a significant p-value always lower than 0.001 (Table 3).

Step	Candidate Gene	Wilks' Lambda	Exact F	p-value
1	TNFAIP3	0.404	59.002	<0.001
2	IL-21	0.300	45.521	<0.001
3	c-REL	0.261	35.809	<0.001
4	RGS1	0.235	30.143	<0.001
5	LPP	0.222	25.272	<0.001

Table 3. Results of Discriminant analysis in the small intestinal tissue: five genes significantly contribute to the discriminant analysis by a stepwise process. It means that the expression values in the intestinal tissue of TNFAIP3, IL-21, c-REL, RGS1 and LPP genes, it is possible to discriminate the celiac type from the control one.

By multiplying the unstandardized canonical discriminant coefficients to the actual values of the expression of the genes listed in the table, a discriminant score is obtained for each individual, which allows the classification into the diagnostic groups.

$$D - Score = (TNFAIP3 * 0.404) + (IL - 21 * 0.300) + (c - REL * 0.261) + (RGS1 * 0.235) + (LPP * 0.222) + constant$$

By applying the discriminant equation to cases and control a classification table was produced (Table 4) which shows the capacity of the equation to discriminate celiacs from controls, in fact 92.9% of original cases were correctly classified (90% of controls and 95.9% of celiacs).

		Predicted Group Membership		
		Control	Celiac	
Real Group Membership	Control	19 (95%)	1 (5%)	20
	Celiac	2 (9.1%)	20 (90.9%)	22
		21	21	42

Table 4. Classification results of discriminant analysis: by applying the results of Wilks' Lambda discrimination analysis, a classification table was produced and the results compared with the real group membership of the considered samples. As

described, 92.9% of samples (90% of controls and 95.9% of celiacs) were correctly classified.

#### *Discriminant analysis in peripheral blood monocytes*

Encouraged by the results obtained with for duodenal mucosa, we developed a similar linear discriminant analysis for the gene expression in peripheral blood monocytes. We randomly divided our celiac and control cases into two balanced groups. By stepwise statistic entered the expression of four candidate genes were selected by multivariate analysis of CD, similar to the results obtained in the duodenal tissue. LPP, c-REL, KIAA1109 and TNFAIP3 well discriminated between cases and controls, and the discriminant function revealed a very low Wilks' lambda (0.048). Indeed 91% of controls and all CD patients were correctly predicted.

To verify these results we applied the discriminant equation obtained in the training set, constituted by 9 CD patients and 11 controls, to develop the discriminant equation, to the gene expression of a new cohort of patients ("validation set") constituted by 7 controls, 8 CD, 5 CD-GFD patients and 9 Crohn patients as positive controls of inflammation. The expression RQ of each individual was multiplied by the unstandardized coefficient obtained with the training set. The D-score was obtained as explained previously. We obtained four clustered D-scores, one for each group (Controls, CD, Crohn and CD on gluten free diet, Figure 9) with no overlap among them. The very low Wilk's lambda obtained by the selected genes, closed to 0=complete discrimination, guarantees the confidence into the classification capacity of the equation in the clinical setting. Figure 9 shows the distribution of the D-Scores of CD, and CD patients on GFD, controls and Cohn's disease patients. The D-Score for active celiac patients was negative in all cases, while it was positive for all the other groups on differentiated clusters.

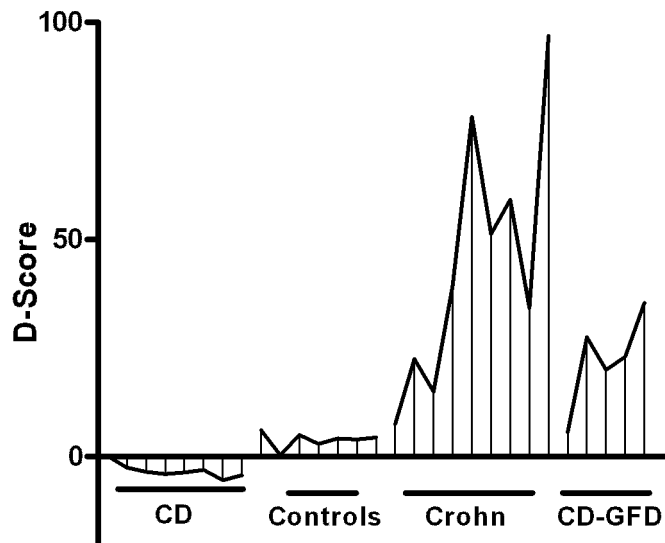


Figure 9. Distribution of the Discriminant of CD, Controls, Crohn and CD-GFD patients. The D-score clearly separated the four groups of subjects evaluated. Only CD patients had a negative D-score. The D-score of CD-GFD patients was intermediate between the scores of controls and Crohn patients.

This score allows indeed to compute a group membership probability for each individual, we were allowed so to classify correctly all controls and CD patients; none of the controls, neither CD on GFD nor Crohn patients were mis-classified as CD patients (Table 5).

Case Number	Actual Group	Highest Group					Discriminant Score
		Predicted Group	P(D>d   G=g)		P(G=g   D=d)	Squared Mahalanobis Distance to Centroid	Function 1
			p	df			
1	Celiac	Celiac	0.000	1	0.878	14.940	0.076
2	Celiac	Celiac	0.222	1	1.000	1.491	-2.568
3	Celiac	Celiac	0.774	1	1.000	0.082	-3.503
4	Celiac	Celiac	0.844	1	1.000	0.038	-3.986
5	Celiac	Celiac	0.882	1	1.000	0.022	-3.641
6	Celiac	Celiac	0.463	1	1.000	0.538	-3.056
7	Celiac	Celiac	0.101	1	1.000	2.695	-5.431
8	Celiac	Celiac	0.581	1	1.000	0.305	-4.341
9	Control	Control	0.089	1	1.000	2.884	6.119
10	Control	Control	0.000	1	0.576	16.548	0.353
11	Control	Control	0.566	1	1.000	0.330	4.995
12	Control	Control	0.126	1	1.000	2.346	2.889
13	Control	Control	0.826	1	1.000	0.048	4.201
14	Control	Control	0.642	1	1.000	0.216	3.956
15	Control	Control	0.955	1	1.000	0.003	4.364
16	Crohn	Control	0.002	1	1.000	9.216	7.457
17	Crohn	Control	0.000	1	1.000	326.316	22.485
18	Crohn	Control	0.000	1	1.000	111.285	14.970
19	Crohn	Control	0.000	1	1.000	1232.648	39.530
20	Crohn	Control	0.000	1	1.000	5440.147	78.178
21	Crohn	Control	0.000	1	1.000	2193.662	51.257
22	Crohn	Control	0.000	1	1.000	2997.302	59.169
23	Crohn	Control	0.000	1	1.000	885.652	34.181
24	Crohn	Control	0.000	1	1.000	8543.655	96.853
25	CD-GFD	Control	0.232	1	1.000	1.431	5.617
26	CD-GFD	Control	0.000	1	1.000	532.145	27.489
27	CD-GFD	Control	0.000	1	1.000	241.667	19.967
28	CD-GFD	Control	0.000	1	1.000	344.648	22.986
29	CD-GFD	Control	0.000	1	1.000	959.294	35.393

Table 5. Validation of the discriminant analysis. The analysis conducted to classify controls, CD, Crohn and CD-GFD patients in 2 groups: Controls and Celiacs. The Highest Group corresponds to the first prediction choice, and the Second Highest Group to the second one. Effective control and celiac patients were all correctly predicted; Crohn and CD-GFD patients were predicted as controls.



## Discussion

Our previous works on genetic risk models and potential CD (Izzo et al., 2011; Sperandeo et al., 2011) reinforce the selection of a small set of candidate genes with specific functions most likely related to the pathogenesis of CD. Indeed the main reason why gene expression studies rarely produce significant clinical predictions is the overwhelming abundance of data. Hundreds of gene over- or under-expressed in the disease status compared to the healthy status very often reflect the inflammation and the expected immune response, but rarely point to the pathogenic profile of the disease.

Thanks to the recently proposed new ESPGHAN diagnostic algorithm for CD, (Husby et al., 2012) duodenal biopsy can now be avoided in a sizeable proportion of patients on the basis of clinical symptoms and high production of anti-tTG antibody. Unfortunately, it is not possible to generalize the method because not all CD patients have high anti-tTG levels (*i.e.* Potential Celiac Disease), and many patients are asymptomatic. Moreover, this new protocol is not applicable for at-risk relatives (Husby et al., 2012).

By virtue of this, the aim of this study was: firstly, to build a panoramic view on the gene expression profile of some of the most associated CD related genes, exploring it at different levels (*i.e.* mucosal biopsies, peripheral monocytes and TCLs). Secondly, to use the gene expression profile of CD-associated genes in peripheral blood monocytes to differentiate patients affected by CD from controls as a step towards the molecular diagnosis of the disease.

First of all we demonstrated that some gene expression profile are modified according to the healthy state of the patients (*e.g.* Differences between gene expression in celiacs and not celiacs patients), in several

cases it was ascribable to the inflammation state of the analyzed tissue and/or cell lines analyzed (comparing gene expression on monocytes of CD and Crohn patients used as positive inflammation controls), in many cases it has been possible to hypothesize to assign the perturbation state to the CD genetic state, because of the similar phenotype between celiacs and CD-GFD patients.

It is noteworthy that KIAA1109 gene expression is over activated in mucosal biopsies of CD-GFD patients and that in these patients it remains higher than controls also at peripheral level, and comparable with that of CD and Crohn. In this case it is very hard to hypothesize the role of this (probable) endonuclease, what is clear is that the same behavior that we checked for CD and Crohn in monocytes suggest us a reasonable mechanism inflammation but not gliadin exposure dependent.

Conversely, for c-REL, SH2B3 and RGS1 at peripheral level we found a different gene expression in CD compared with both controls and Crohn. These findings are important mostly for c-REL genes, because we demonstrated that, by virtue of the involvement of this protein in NF- $\kappa$ B pathway, the inflammation impairment occurs not only at tissue target but also at peripheral level. Regarding c-REL and SH2B3 we did confirm an alteration also at intestinal mucosa level; conversely this is not true for RGS1 gene, which expression is particularly altered in mucosa-derived TCLs. It is important to highlight these findings because RGS1 gene encodes a member of G protein signaling regulator family which is probably involved in homing mechanism. In fact *Rgs1*<sup>-/-</sup> mice showed an increased B cell movement from and towards lymph nodes (Han et al., 2005) and an increased migratory response of dendritic cells after chemokines stimulation (Shi et al., 2004). Interestingly in our studies, RGS1 gene is down-regulated at basal level in celiac T cell clones, and its levels become higher than controls after 3h-incubation with gliadin

peptides but not with medium only: these findings demonstrate that (i) RGS1 is directly regulated by gliadin peptides; (ii) CD patients have an altered homing mechanism which should be deeper investigated.

Surprising, very positive results were from gene expression on peripheral monocytes, so the second step of our work has been to use the information provided from gene expression profile to build a prognostic (and hopefully diagnostic) equation able to discriminate celiac type from the control one. Using a discriminant analysis, here we demonstrate that the expression of a small set of candidate genes in peripheral blood monocytes can be used to correctly classify CD patients from controls and from patients affected by Crohn disease, without considering HLA and anti-tTG levels. In fact, the procedure we used resulted in a distance between groups that is unusual with ordinary diagnostic tools (very low Wilk's lambda). We did not add anti-tTG antibodies or HLA data for each individual because it is well recognized that the former has a very high sensitivity and specificity, and the latter has a very high negative predictive value. Since we reached a correct diagnostic classification in the validation set (above 95%) by gene expression data only, we would have shown overoptimistic estimate by adding the anti-tTG values. But, this should be done in clinical practice in order to increase the sensitivity and the specificity of the diagnosis when duodenal biopsy is not available.

## **Chapter 2**

**Non-HLA genes and Celiac Disease:**

**Potential CD patients**

## Chapter 2. Non-HLA genes and Celiac Disease: Potential CD patients

### Background and aim

Recently the clinical presentation of CD underwent substantial changes: from classical malabsorption to light or absent symptoms associated to moderate damage of intestinal mucosa. The term '*potential CD*' was assigned to individuals with proper DQ2 or DQ8, production of anti tTG, normal small intestinal mucosa (Figure 10), classified as Marsh 0 (no damage) or Marsh 1 stage (unspecific intra epithelial infiltration only). These are also defined "serological positive cases" (Koskinen et al., 2008; Kurppa et al., 2010). Potential patients suggest that the development of adaptive anti-gluten immunity is not sufficient to develop villous atrophy. This is supported by animal models showing that inflammatory gluten-induced responses are not sufficient to develop mucosal lesions. Thus, we can subgroup individuals into three groups: controls, patients who developed antibodies but no intestinal lesions (M0 and M1 potentials) and patients with antibodies and villous atrophy (overt CD). Analyzing differences among these subgroups may shed light into the genes differentially involved in the development of adaptive gluten-induced immunity and of the tissue lesions (different between potential and CD), as well as genes involved in all steps of CD pathogenesis (different between control and potential, and potential and CD).

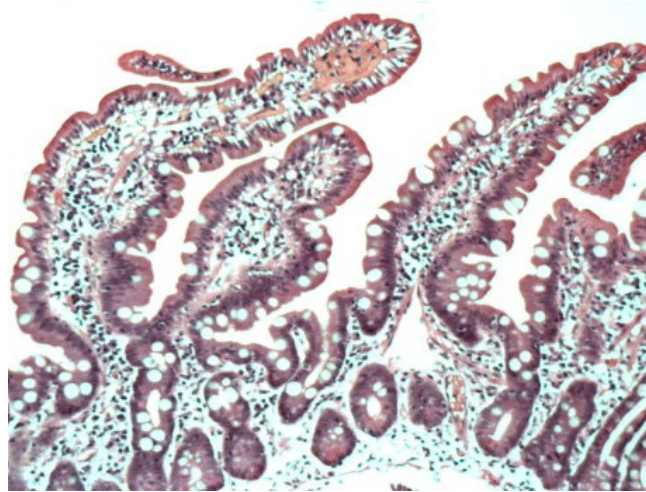


Figure 10: duodenal mucosa of a Potential CD patient.

We have a sizeable cohort of a living experimental model of CD: intestinal production of anti-tTG antibodies but no specific mucosal damage. The majority of these individuals have no symptoms and may be grouped into low-medium-high anti-tTG producers, but all have a small intestinal mucosa with normal villi/crypt ratio, none or moderate IEL infiltration and normal epithelial layer. Over a prolonged follow up on gluten containing diet only about 1/3<sup>rd</sup> develop mucosal damage, most frequently not associated to symptoms (Tosco et al., 2011). The reasons why potential CD do not show any degree of substantial mucosal damage, albeit the presence of all the features of CD, is the question, whose answer can contribute to the understanding of the pathogenesis of CD.

Our aim was to explore the presence of genetic and expression factors that may differentiate potential patients from overt CD patients and controls, in particular:

- To find a different distribution of HLA haplotypes;
- To evaluate if there is a different genotypic distribution of non-HLA;

- To confirm association results also in expression studies on RNA extracted from intestinal biopsies.
- To understand if a correlation exists between genotype and expression levels.

A further step was to evaluate if these analysis could be applied in order to compare potentials CD cases which developed and those which not developed villous atrophy after a three years-follow up program (Tosco et al., 2011).

## **Methods**

### ***Subjects***

1. 127 patients from southern Italy (median age 6 years and 6 months, range 18 months-16 years) were classified as 'potential CD cases' on the basis of:
  - Increased levels of anti-tTG (IgA anti Human Tissues Transglutaminase) and Anti-Endomisium positive. Serum EMA and anti tTG IgA were detected by indirect immunofluorescence and by enzyme-linked immunosorbent assay (ELISA) technique using a kit based on human recombinant antigen respectively (Eu-tTg IgA Kit, Eurospital, Trieste).
  - Normal architecture of the small intestinal mucosa (Marsh 0 and Marsh 1 stages) on at least four mucosa samples, to minimize the possible bias due to patchy lesions. Definition of the stage of lesion was obtained by Marsh stages modified by Oberhuber (Tosco et al., 2011).

After a three years follow-up, 11 potential CD patients developed villous atrophy.

2. To evaluate the HLA related risk haplotypes, the 127 potential cases were compared to a set of 311 overt cases from the same population described above and reported in Bourgey et al (Bourgey et al., 2007).
3. To evaluate non-HLA related risk alleles, a cohort of 643 overt CD cases, and 711 controls were compared to the 127 potential CD patients: in particular, genotyping of the candidate genes on 4q27 locus was performed on a set of about 350 cases and about 450 controls from the above sample. All tested cases originated from the same centre and the same geographical area in order to minimize the bias due to the population stratification. Controls were randomly selected from a DNA bank representing the healthy population of the region.

### ***Genotyping***

HLA typing was performed using Eu-Gene Risk kit (EUROSPITAL, Trieste, Italy). Cases were grouped into 5 HLA (double DQ2, DQ2 in trans, DQ2 single, DQ8 + DQB1\*02, No DQ2or DQ8), as previously reported (Bourgey et al., 2007; Margaritte-Jeannin et al., 2004).

Thirteen SNPs (rs6441961, rs17810546, rs1738074, rs917997, rs2816316, rs1464510, rs2327832, rs842647, rs6822844, rs4374642, rs13119723, rs1127348, rs6840978) were genotyped using TaqMan technology (Life Technologies, Foster City, CA) (Table 6). Reactions were performed on 7900HT Fast Real-Time PCR System (Life Technologies, Foster City, CA) in a volume of 20 µL, containing SNP master mix, TaqMan



assays and about 50 ng of genomic DNA template. Each plate also contained genotyping controls (4 duplicates of 4 CEPH DNA).

Chr	Gene	SNP
1	<b>RGS1</b>	rs2816316
2	<b>IL18RAP</b>	rs917997
2	<b>REL</b>	rs842647
3	<b>SCHIP</b>	rs17810546
3	<b>LPP</b>	rs1464510
3	<b>CCR</b>	rs6441961
4	<b>KIAA1109</b>	rs1127348
4		rs13119723
4		rs4374642
4	<b>IL2</b>	rs6822844
4	<b>IL21</b>	rs6840978
6	<b>OLIG3</b>	rs2327832
6	<b>TAGAP</b>	rs1738074

Table 6. SNPs analyzed in our cohort

### ***Statistical Analysis***

Frequencies were compared by the  $\chi^2$  test with 0.05 probability of the null hypothesis. Skewed variables were log transformed when required. Expression data were compared by a signed Rank test for independent samples. Data analysis was performed using: Prism (GraphPad Software, San Diego, CA), SPSS 15 (SPSS Inc., Chicago, IL, USA) and Haploview 4.1.

## **Results**

### ***HLA typing***

The HLA genotype of 127 potential CD cases was compared to that of 311 Celiacs previously reported by us (Bourgey et al., 2007) (Table 7 and 8).

HLA	Haplotype DQA1-DQB1	Linkage DR Allele
DQ2.5	DQA1*05-DQB1*0201	DR3
½ DQ2 <i>in trans</i>	DQA1*02-DQB1*0202	DR7
½ DQ2 <i>in cis</i>	DQA1*05-DQB1*03	DR5
DQ8	DQA1*0301-DQB1*0302	DR4
DQX	ALTRE	DRX

Table 7. HLA haplotypes associated with CD.

DQ	DR in linkage	HLA risk Group
DQ2.5/DQ2.5	DR3/DR3	Group 1
DQ2.5/DQ2.2	DR3/DR7	
DQ7/DQ2.2	DR5/DR7	Group 2
DQ2.5/DQ7	DR3/DR5	Group 3
DQ2.5/DQ8	DR3/DR4	
DQ2.5/DQX	DR3/DRX*	
DQ2.2/DQ2.2	DR7/DR7	Group 4
DQ2.2/DQ8	DR7/DR4	
DQ8/DQ8	DR4/DR4	
DQ2.2/DQX	DR7/DRX*	
DQ7/DQ7	DR5/DR5	Group 5
DQ7/DQX	DR5/DRX*	
(more frequent)		
X* different from 3, 4, 5, 7		

Table 8. Classification according to the HLA genotype (Bourgey et al., 2007). DQ2.5 → DQA1\*05 – DQB1\*0201; DQ2.2 → DQA1\*02 – DQB1\*0202; DQ7 → DQA1\*05 – DQB1\*03; DQ8 → DQA1\*0301 – DQB1\*0302

Potential CD less frequently belong to high risk classes, while they show more frequently very low or moderate HLA-related risk (Figure 11). We found a significant HLA distribution in Potential CD cases versus celiacs ( $\chi^2=48.42$   $p<0.001$ ). When we compared the HLA risk between M0 and M1 potential cases we did not observe any difference ( $\chi^2 =1.92$ ;  $p=0.75$ , data not shown). Cases who developed small intestinal atrophy during follow up did not show an HLA distribution different from those who remained potential ( $\chi^2 =5.17$ ;  $p=0.27$ , data not shown).

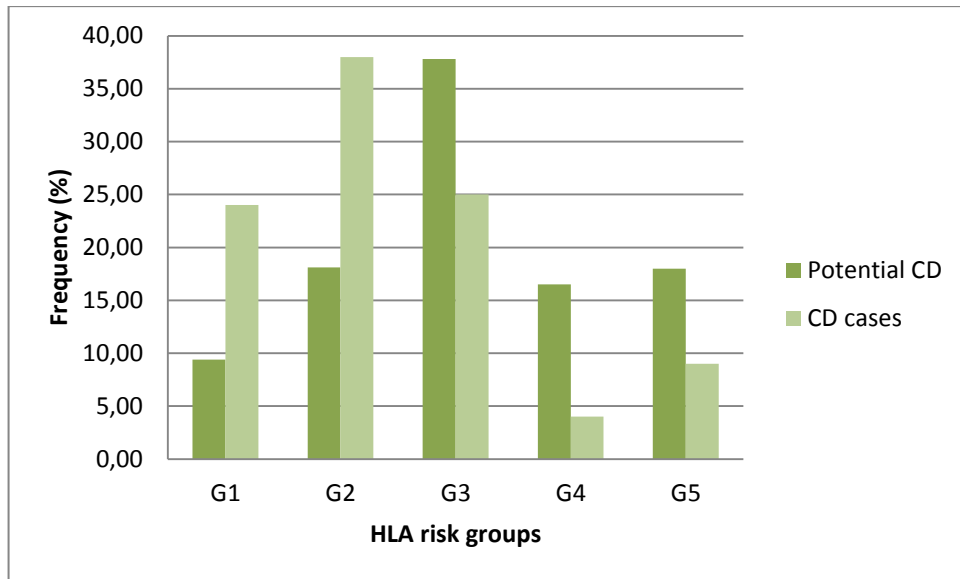


Figure 11. Distribution of HLA haplotypes in Potential CD and CD cases according to the risk groups described in Bourgey et al, 2007 (Bourgey et al., 2007).

### ***Eight candidate genes SNPs genotype and allele distribution***

Hunt et al. (Hunt et al., 2008) reported 8 new CD risk loci, located on different chromosomes. Although we know that the size of the sample of the potential CD has no sufficient power to differentiate the polymorphism's distribution, we found significant results in the genotype and allele distribution when we compared Potential CD *versus* controls and CD cases.

From a genotypic point of view, Potential CD show a different distribution of the SNPs located in c-REL and CCR genes *versus* control; seven out of 8 SNPs, showed no significant differences between potential and CD cases in their genotype distribution, but there was a significant difference of the SNP in c-REL gene also in this case (Table 9).

		Controls	Potential CD	$\chi^2$	<i>p</i> value	CD cases	Potential CD	$\chi^2$	<i>p</i> value
<b>RGS1</b> <b>(rs2816316)</b>	AA	488 (68.6%)	88 (69.3%)	0.32	0.85	463 (72.7%)	88 (69.3%)	2.67	0.26
	AC	193 (27.1%)	35 (27.6%)			166 (26.1%)	35 (27.6%)		
	CC	30 (4.2%)	4 (3.1%)			8 (1.3%)	4 (3.1%)		
<b>IL18RAP</b> <b>(rs917997)</b>	AA	42 (5.9%)	7 (5.5%)	1.07	0.58	41 (6.4%)	7 (5.5%)	0.69	0.71
	AG	246 (34.6%)	50 (39.4%)			227 (35.6%)	50 (39.4%)		
	GG	423 (59.5%)	70 (55.1%)			369 (57.9%)	70 (55.1%)		
<b>LPP</b> <b>(rs1464510)</b>	AA	108 (15.2%)	28 (22.0%)	<b>7.58</b>	<b>0.02</b>	152 (23.9%)	28 (22.0%)	0.77	0.68
	AC	362 (50.9%)	70 (55.1%)			324 (50.9%)	70 (55.1%)		
	CC	241 (33.9%)	29 (22.8%)			161 (25.3%)	29 (22.8%)		
<b>OLIG3</b> <b>(rs2327832)</b>	AA	480 (67.7%)	84 (66.7%)	0.05	0.98	399 (63.6%)	84 (66.7%)	0.81	0.67
	AG	202 (28.5%)	37 (29.4%)			208 (33.2%)	37 (29.4%)		
	GG	27 (3.8%)	5 (4.0%)			20 (3.2%)	5 (4.0%)		
<b>TAGAP</b> <b>(rs1738074)</b>	AA	125 (17.6%)	26 (20.5%)	0.6	0.74	144 (22.6%)	26 (20.5%)	0.36	0.84
	AG	354 (49.9%)	61 (48.0%)			305 (47.9%)	61 (48.0%)		
	GG	231 (32.5%)	40 (31.5%)			188 (29.5%)	40 (31.5%)		
<b>c-REL</b> <b>(rs842647)</b>	AA	404 (56.8%)	62 (48.8%)	<b>9.84</b>	<b>0.01</b>	359 (56.4%)	62 (48.8%)	<b>10.8</b>	<b>0.01</b>
	AG	272 (38.3%)	50 (39.4%)			249 (39.1%)	50 (39.4%)		
	GG	35 (4.9%)	15 (11.8%)			29 (4.6%)	15 (11.8%)		
<b>CCR</b> <b>(rs6441961)</b>	AA	99 (13.9%)	25 (19.7%)	<b>8.28</b>	<b>&lt;0.01</b>	120 (18.8%)	25 (19.7%)	2.29	0.32
	AG	318 (44.7%)	66 (52.0%)			293 (46.0%)	66 (52.0%)		
	GG	294 (41.4%)	36 (28.3%)			224 (35.2%)	36 (28.3%)		
<b>SCHIP1</b> <b>(rs17810546)</b>	AA	610 (85.8%)	102 (80.3%)	2.56	0.28	526 (82.6%)	102 (80.3%)	0.92	0.63
	AG	92 (12.9%)	23 (18.1%)			106 (16.6%)	23 (18.1%)		
	GG	9 (1.3%)	2 (1.6%)			5 (0.8%)	2 (1.6%)		

Table 9. Genotypic distribution of 8 non-HLA SNPs in Potential CD versus controls and celiac cases.

By the analysis of the allele distribution, we found that both LPP\*A, c-REL\*G, and CCR\*A alleles were significantly associated to potential cases compared to controls (Table 10). Only c-REL\*G allele was significantly associated with potential cases when compared to CD cases (Table 10). Although RGS1 and CCR produced an interesting difference (in term of  $\chi^2$  which is 1.31 and 1.61 respectively), they could not reach the statistical significance (Table 10).

Potential CD and controls									
SNP	Locus	A1	A2	Pot CD Cases, Controls Ratios	MAF Pot CD cases	MAF controls	$\chi^2$	p value	Odds ratio (95% confidence interval)
rs2816316	RGS1	A	C	214:44, 1169:253	0.17	0.18	0.08	0.77	1.05 (0.74-1.50)
rs917997	IL18RAP	G	A	193:65, 1092:330	0.25	0.23	0.48	0.49	0.90 (0.66-1.22)
<b>rs1464510</b>	<b>LPP</b>	<b>C</b>	<b>A</b>	<b>129:129, 844:578</b>	<b>0.50</b>	<b>0.41</b>	<b>7.84</b>	<b>0.01</b>	<b>0.68 (0.52-0.89)</b>
rs2327832	OLIG3	A	G	207:49, 1162:256	0.19	0.18	0.17	0.68	0.93 (0.66-1.31)
rs1738074	TAGAP	G	A	143:115, 816:604	0.44	0.42	0.37	0.54	0.92 (0.70-1.20)
<b>rs842647</b>	<b>c-REL</b>	<b>A</b>	<b>G</b>	<b>178:80, 1080:342</b>	<b>0.31</b>	<b>0.24</b>	<b>5.62</b>	<b>0.02</b>	<b>0.70 (0.53-0.94)</b>
<b>rs6441961</b>	<b>CCR</b>	<b>G</b>	<b>A</b>	<b>139:119, 906:516</b>	<b>0.46</b>	<b>0.36</b>	<b>8.99</b>	<b>&lt;0.01</b>	<b>0.66 (0.51-0.87)</b>
rs17810546	SCHIP1	A	G	231:27, 1312:110	0.10	0.08	2.17	0.14	0.72 (0.46-1.12)
Potential CD and CD cases									
SNP	Locus	A1	A2	Pot CD cases. CD cases Ratios	MAF Pot CD cases	MAF CD cases	$\chi^2$	p value	Odds ratio (95% confidence interval)
rs2816316	RGS1	A	C	214:44, 1092:182	0.17	0.14	1.31	0.25	0.81 (0.56-1.16)
rs917997	IL18RAP	G	A	193:65, 965:309	0.25	0.24	0.10	0.75	0.95 (0.70-1.29)
rs1464510	LPP	C	A	129:129, 646:628	0.50	0.49	0.04	0.83	0.97 (0.74-1.27)
rs2327832	OLIG3	A	G	207:49, 1008:248	0.19	0.20	0.05	0.82	1.04 (0.74-1.46)
rs1738074	TAGAP	G	A	143:115, 681:593	0.44	0.46	0.34	0.56	1.08 (0.83-1.42)
<b>rs842647</b>	<b>c-REL</b>	<b>A</b>	<b>G</b>	<b>178:80, 967:307</b>	<b>0.31</b>	<b>0.24</b>	<b>5.43</b>	<b>0.02</b>	<b>0.71 (0.53-0.95)</b>
rs6441961	CCR	G	A	139:119, 741:533	0.46	0.42	1.61	0.20	0.84 (0.64-1.10)
rs17810546	SCHIP1	A	G	231:27, 1158:116	0.10	0.09	0.47	0.49	0.84 (0.55-1.33)

Table 10. Allelic distribution of 8 non-HLA SNPs in Potential CD versus controls and celiac cases.

None of the candidate genes distinguished between M0 and M1 potential cases. M0 Potential CD cases were not different from controls, except for the CCR gene. M1 patients reinforced the significant differential distribution of the c-REL genotype when compared to CD: again the GG genotype was about three times more frequent in M1 patients compared to CD patients. The same trend, with good statistical significance, is

observed between M1 potential and controls not only for c-REL but also for LPP gene (Table 11).

		Controls	Potential M0	Potential M1	CD cases	M0/Controls		M1/CD		M1/Controls	
						$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>
<b>c-REL</b> <b>(rs842647)</b>	<b>AA</b>	404 (56.8%)	26 (47.3%)	33 (47.8%)	359 (56.4%)	<b>2.91</b>	<b>0.23</b>	<b>11.98</b>	<b>&lt;0.01</b>	<b>10.9</b>	<b>&lt;0.01</b>
	<b>AG</b>	272 (38.3%)	24 (43.6%)	26 (37.7%)	249 (39.1%)						
	<b>GG</b>	35 (4.9%)	5 (9.1%)	10 (14.5%)	29 (4.6%)						
<b>CCR</b> <b>(rs6441961)</b>	<b>AA</b>	99 (13.9%)	11 (23.4%)	9 (15.8%)	120 (18.8%)	<b>6.12</b>	<b>0.04</b>	1.20	0.55	4.22	0.12
	<b>AG</b>	318 (44.7%)	24 (51.1%)	30 (52.6%)	293 (46.0%)						
	<b>GG</b>	294 (41.4%)	12 (25.5%)	18 (31.6%)	224 (35.2%)						
<b>LPP</b> <b>(rs1464510)</b>	<b>AA</b>	108 (15.2%)	12 (21.8%)	15 (21.7%)	152 (23.9%)	<b>2.10</b>	<b>0.35</b>	1.35	0.51	<b>5.88</b>	<b>0.05</b>
	<b>AC</b>	362 (50.9%)	28 (50.9%)	40 (58.0%)	324 (50.9%)						
	<b>CC</b>	241 (33.9%)	15 (27.3%)	14 (20.3%)	161 (25.3%)						

Table 11. Comparison between cases with completely normal mucosa (M0) and those with infiltrated mucosa (M1).

The eleven cases which developed small intestinal atrophy during follow up did not show an allelic and genotypic distribution different from those who remained potential, maybe because of the sample size that was too small to evaluate differences.

### ***Five Candidate SNPs on the 4q genomic region***

The 4q27 region, among non-HLA genes, is the one resulted to be associated with CD in all GWAs and replications performed until now.

Despite the limited sample size we found that three SNPs in the 4q genomic region showed significant differences between controls and potential CD patients (rs4374642, rs13119723 and rs6840978) in their genotype distribution. In addition, we found also significant differences between potential CD and CD patients for three SNPs (rs4374642, rs13119723 and rs6822844) (Table 12).

		Controls	Potential CD	$\chi^2$	<i>p</i> value	CD cases	Potential CD	$\chi^2$	<i>p</i> value
<b>KIAA1109 (rs4374642)</b>	CC					4 (1.2%)			
	CT	63 (14.2%)	9 (7.1%)	<b>4.52</b>	<b>0.03</b>	53 (15.3%)	9 (7.1%)	<b>7.22</b>	<b>0.03</b>
	TT	381 (85.8%)	118 (92.9%)			290 (83.6%)	118 (92.9%)		
<b>KIAA1109 (rs13119723)</b>	AA	397 (88.6%)	104 (81.3%)	<b>5.99</b>	<b>0.05</b>	290 (84.1%)	104 (81.3%)	<b>8.16</b>	<b>0.02</b>
	AG	48 (10.7%)	21 (16.4%)			55 (15.9%)	21 (16.4%)		
	GG	3 (0.7%)	3 (2.3%)				3 (2.3%)		
<b>KIAA1109 (rs1127348)</b>	CC	16 (3.6%)	8 (6.3%)	2.05	0.36	9 (2.6%)	8 (6.3%)	3.51	0.17
	CT	122 (27.5%)	37 (29.1%)			103 (30.3%)	37 (29.1%)		
	TT	205 (68.8%)	82 (64.6%)			228 (67.1%)	82 (64.6%)		
<b>IL2/IL21 (rs6822844)</b>	GG	358 (80.1%)	103 (79.8%)	0.35	0.84	267 (81.9%)	103 (79.8%)	<b>7.61</b>	<b>0.02</b>
	GT	82 (18.3%)	23 (17.8%)			59 (18.1%)	23 (17.8%)		
	TT	7 (1.6%)	3 (2.3%)				3 (2.3%)		
<b>IL21 (rs6840978)</b>	CC	363 (81.6%)	85 (72.0%)	<b>17.94</b>	<b>&lt;0.01</b>	222 (76.8%)	85 (72.0%)	4.5	0.10
	CT	82 (18.4%)	29 (24.6%)			65 (22.5%)	29 (24.6%)		
	TT		4 (3.4%)			2 (0.7%)	4 (3.4%)		

Table 12. Association results for 5 celiac non-HLA risk variants, located on 4q27 locus.

Although rs1127348 and rs6840978 produced an interesting difference between CD cases and Potential CD (in term of  $\chi^2$  which is 3.51 and 4.5 respectively), they could not reach the statistical significance (Table 12).

The rs4374642, rs13119723 and rs6840978 SNPs, respectively, were significantly associated when Potential CD cases were compared to Controls. In contrast, only the rs4374642\*T was the risk allele associated with potential cases when compared to CD cases (Table 13).

Potential CD and Controls

SNP	Locus	A1	A2	Pot CD Cases, Controls Ratios	MAF Pot CD cases	MAF Controls	$\chi^2$	<i>p</i> value	Odds ratio (95% confidence interval)
rs4374642	KIAA1109	T	C	249:9, 823:63	0.03	0.07	4.44	0.03	2.12 (1.04-4.32)
rs13119723	KIAA1109	A	G	229:27, 832:54	0.10	0.06	5.97	0.01	0.55 (0.34-0.89)
rs1127348	KIAA1109	T	C	201:53, 732:154	0.21	0.17	1.61	0.20	0.80 (0.56-1.13)
rs6822844	IL2/IL21	G	T	228:28, 796:96	0.11	0.10	0.01	0.94	0.98 (0.63-1.53)
rs6840978	IL21	C	T	199:37, 802:82	0.16	0.09	8.04	<0.01	0.55 (0.36-0.83)

Potential CD and CD patients

SNP	Locus	A1	A2	Pot CD Cases, CD cases Ratios	MAF Pot CD cases	MAF CD cases	$\chi^2$	<i>p</i> value	Odds ratio (95% confidence interval)
rs4374642	KIAA1109	T	C	249:9, 622:58	0.03	0.08	7.17	0.01	2.58 (1.26-5.29)
rs13119723	KIAA1109	A	G	229:27, 604:54	0.10	0.08	1.25	0.26	0.76 (0.47-1.23)
rs1127348	KIAA1109	T	C	201:53, 552:120	0.21	0.18	1.10	0.29	0.82 (0.58-1.18)
rs6822844	IL2/IL21	G	T	228:28, 592:58	0.11	0.09	0.87	0.35	0.79 (0.49-1.28)
rs6840978	IL21	C	T	199:37, 503:69	0.16	0.12	1.92	0.17	0.74 (0.48-1.14)

Table 13. Association results of non-HLA SNPs Alleles, located on chromosome 4q27.

Cases which developed small intestinal atrophy during follow up did not show an allelic (Table 14) and genotypic (data not shown) distribution different from those which remained potentials.



Potential CD which develop villous atrophy and still Potential CD cases

SNP	Locus	A1	A2	Pot CD with atrophy, Still Pot Ratios	MAF Pot CD with atrophy	MAF Still Potential CD	$\chi^2$	p value	Odds ratio (95% confidence interval)
rs2816316	RGS1	A	C	16:6, 195:37	0.27	0.15	1.83	0.17	0.51 (0.18-1.38)
rs917997	IL18RAP	G	A	18:4, 172:60	0.18	0.25	0.63	0.43	1.57 (0.51-4.82)
rs1464510	LPP	A	C	12:10, 114:118	0.45	0.49	0.23	0.63	1.24 (0.52-2.99)
<b>rs2327832</b>	<b>OLIG3</b>	<b>A</b>	<b>G</b>	<b>14:8, 191:39</b>	<b>0.36</b>	<b>0.16</b>	<b>4.98</b>	<b>0.02</b>	<b>0.36 (0.14-0.91)</b>
rs1738074	TAGAP	G	A	13:9, 128:104	0.40	0.44	0.12	0.72	1.17 (0.48-2.85)
rs842647	c-REL	A	G	15:7, 159:73	0.31	0.31	0.01	0.97	0.98 (0.38-2.52)
rs6441961	CCR	G	A	14:8, 124:108	0.36	0.46	0.84	0.36	1.52 (0.61-3.78)
rs17810546	SCHIP1	A	G	20:2, 207:25	0.09	0.10	0.06	0.80	1.21 (0.27-5.48)
rs4374642	KIAA1109	T	C	22:0, 219:9	0.00	0.03	0.90	0.34	1.95 (1.10-34.61)
rs13119723	KIAA1109	A	G	21:1, 204:26	0.04	0.11	0.96	0.33	2.67 (0.34-20.74)
rs1127348	KIAA1109	T	C	18:4, 179:49	0.18	0.21	0.13	0.72	1.23 (0.40-3.81)
rs6822844	IL2/IL21	G	T	21:1, 204:26	0.04	0.11	0.96	0.33	2.68 (0.34-20.74)
rs6840978	IL21	C	T	14:2, 182:34	0.12	0.14	0.12	0.73	1.31 (0.28-6.02)

Table 14. Association results of 13 non-HLA SNPs alleles.

## Discussion

This work attempts to explore the genetic profiles of potential Celiac patients, characterized by a positive serology and undamaged small intestinal mucosa: they appear to be associated to different genetic risk factors compared with CD patients and controls.

Potential CD have a HLA-related CD risk significantly lighter than that of CD cases; no difference in HLA profile was observed between M0 and M1 potential patients, neither between the cases who developed full blown disease and those who remained potential after a three years follow-up. It's important to note that potential patients without the classical DQ2 or DQ8 heterodimers (HLA risk group 5) were double in potential (18.0%) than in classical celiacs (9.0%), but they produced antibodies against Human anti-tTG, with positive Endomysial antibodies, in some cases as much as CD cases with proper DQ2.5 heterodimer. In addition we could find no discrimination on the outcome between the DQ2/DQ8 carriers and those who carried only half of the heterodimers.

We identified six polymorphisms suggestive to be differently distributed between controls and potential CD: in particular those in KIAA1109, IL-21, LPP, c-REL and CCR genes. We suppose that these factors are implicated at some stages of CD pathogenesis, and that, as for HLA, there may be a "gene-dosage effect".

We are interested especially in c-REL gene, a subunit of the NF- $\kappa$ B complex, crucial for the regulation of this major nuclear factor for the innate and adaptive immunity. Our group produced consistent evidence of the very early involvement of NF- $\kappa$ B mechanism in the gliadin-induced innate immune response in CD (Maiuri et al., 2003): NF- $\kappa$ B is constitutively active in intestinal mucosa of patients with untreated CD and reverts to normal values when gluten is removed from the diet. On these basis we

think that sustained activation of NF- $\kappa$ B in intestinal mucosa of CD patients leads to prolonged induction of inflammatory genes expression and thereby perpetuates the chronic inflammatory process.

Potential patients have most of the features of CD patients: they have similar genes, produce the same antibodies but they lack the final destructive phase of the gluten induced immunoresponse, the cytotoxic destruction of small intestinal mucosa. No clinical, serological or histological features could distinguish M0 from M1 patients up to now (Tosco et al., 2011), for the first time we can speculate that potential CD are a mixture of two populations, with different non-HLA genetics and significant differential expression of some candidate genes.

Those potential cases who developed full blown mucosal damage on follow up, did not show any of the time 0 serologic markers worse than those who remained potential over a prolonged follow up. Similarly no genetic polymorphism, including HLA, could predict the outcome of these potential cases. We observed the same distribution of the HLA as well as of the non-HLA genes in the cases who eventually become full blown celiacs.

The limit of this study is the relatively short follow up time of this cohort: up to now it was not possible to identify any marker (clinical, environmental, serological or genetic) that could help to predict the time of conversion from predisposition to the disease. In conclusion we cannot state that the cohort of potential celiacs who have again shown an undamaged intestinal mucosa after 4 to 6 years of follow up, will not eventually develop it later in life. We can only observe that this group did not develop the full blown disease after 4-6 years. But the genetic and expression features point to the existence of a peculiar molecular profile, that may help to identify cases who do not have the complete molecular repertoire to develop the disease. They may indeed be considered a live

biological model of CD pathogenesis, where the process is, for some time or even forever, interrupted just before the final destructive phase.

## **Chapter 3**

### **Non-HLA genes and Celiac Disease:**

#### **A Family study**

### **Chapter 3. Non-HLA genes and Celiac Disease: A Family study**

#### **Background and aim**

Since several studies showed an elevated prevalence of Celiac Disease (CD) in sibs of CD patients, our group previously evaluated the CD risk in sibs of CD patients (Bourgey et al., 2007). The study included a cohort of 188 Italian families, composed by CD probands, at least one sib and both parents and permitted to classify people in five different group of risk according to HLA genotype (Table 8).

The evidence of a strong genetic component is suggested by a remarkable familiar aggregation: the prevalence of CD is, in fact, 10 times higher in first degree relatives (~10%) than in the whole population (1%) (Bonamico et al., 2006; Bourgey et al., 2007; Mustalahti et al., 2010) and a very high concordance (>80%) is found in monozygotic twins (Greco et al., 2002).

In the field of complex diseases, particularly in CD, great attention is now paid to use the available genetic data to predict the risk of disease in asymptomatic individuals and to support the diagnosis in ambiguous cases. Although it was described that non-HLA genes improve the ability to identify individuals at high risk the increased prediction ability by genetics only seems still modest in the general population (Romanos et al., 2009). However the use of non-HLA genes in the disease risk prediction in CD sibs has not yet been fully explored.

The aim of this study was to develop a genetic risk model for CD using HLA and non-HLA risk alleles. For this study, we analysed families previously recruited (Bourgey et al., 2007) to evaluate the CD risk in sibs of children with CD. In fact, the study includes a cohort of 188 Italian families, composed by a CD proband, at least one sib and both parents. Preliminary

data show that at least one non-HLA risk allele increases the risk of CD. This effect can be added in an additive model to HLA risk.

We aim to significantly improve the identification of subjects at risk to develop CD by modeling the HLA risk with the newly discovered genetic profile in order to improve the accuracy in prediction. When this goal is reached, families with one case will know, with reliable accuracy, the individual risk of a newborn and will be able to anticipate the onset of the clinical phenotype by a simple point-of-care tests. After families the model might well be extended to populations, in the frame of national policies to screen for genetic predispositions to diseases.

## **Methods**

### ***Family Samples***

A cohort of CD families was recruited in a previous study (Bourgey et al., 2007). Families included a symptomatic CD patient (hereafter referred as the proband), both parents and at least one sib with known outcome (for a total of 183 probands, 366 parents and 249 sibs); all probands, as well as the new cases, were diagnosed according to the European Society of Paediatrics Gastroenterology and Nutrition (ESPGHAN) criteria (Husby et al., 2012). Among the 249 sibs, 33 resulted to be affected over a 6 years follow up program.

All individuals were grouped into five decreasing risk classes according to their HLA genotype: very high (>20% with two copies of DQ2.5, or DQ2.5/DQ2.2 Group 1), high (15–20% with DQ2.2/DQA105, Group 2), intermediate (10–15% with one copy of the DQ2.5 heterodimer, Group 3), moderate (1–10% with a double copy of DQ8 or DQ2.2/DQ8, or double

copy of DQ2.2, Group 4) or negligible (<1%, with other genotype, Group 5) (Table 8).

The informed consent was obtained from all participants and from both parents for children. The research was approved by the Ethics Committee of the School of Medicine, University of Naples “Federico II” and was according to principles of the Helsinki II declaration.

### ***Non-HLA Single Nucleotide Polymorphisms (SNPs) typing***

The 798 patients were genotyped for 10 non-HLA SNPs associated with CD: rs6441961 on 3p21 (CCR1/CCR3), rs17810546 and rs9811792 on 3q25-26 (IL12A/SCHIP1 and IL12A), rs1738074 on 6q25 (TAGAP – *T cell activation GTPase activating protein*), rs2816316 on 1q31 (RGS1 – *Regulator of G-protein signaling 1*), rs1464510 on 3q28 (LPP – *Lipoma Preferred Partner*), rs2327832 on 6q23.3 (OLIG3), rs842647 on 2p16.1 (REL), rs6822844 on 4q27 (IL2/IL21), rs3184504 on 12q24 (SH2B3). Genotyping reactions were performed using TaqMan®SNP Genotyping Assays on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA); the final volume was 5 µL, containing master mix, TaqMan assays and about 60 ng of genomic DNA template. All 384 well plates were filled using Biomek® FX (Beckman Coulter, Indianapolis, IN, USA). Allelic Discrimination results were analyzed through the SDS software ver. 2.3.



### ***Analysis strategy and statistics***

In order to develop the model we splitted the sample into a training and a testing set. In the *training set*, composed by probands and their unaffected parents, called *trios*, we evaluated which SNP was associated with the disease, independently from the HLA haplotype. Twenty-six CD families were excluded because they did not meet the requirements for the analysis (e.g. there was an affected parent or patients had an incomplete genotyping), so the training set was composed by 157 trios (Figure 12).

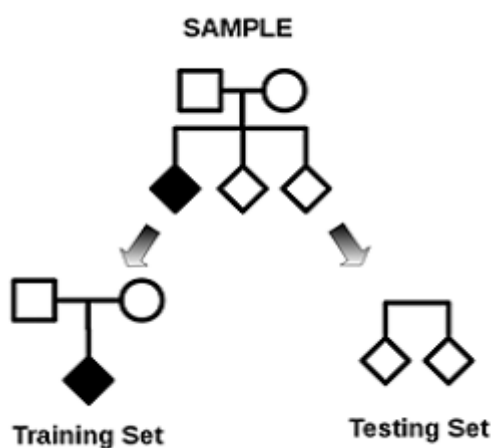


Figure 12. Study design: the family set was splitted in a Testing set – 157 *trios* composed by the 157 probands and their unaffected parents – and a Training set – 252 sibs of the probands.

As control haplotypes we considered the haplotypes carried by parents and not transmitted to the affected probands, as they could be representative of the haplotypes in the general population, from which cases are originated. The frequencies of HLA and non HLA haplotypes in controls were estimated by the AFBAC (Affected Family Based Controls)

method implemented in the MASC software tool (Thomson, 1995). To evaluate the association of the SNPs with the disease, we performed a *Transmission Disequilibrium Test* (TDT) by using the *trio* package ver. 1.1.15 for R statistical computing software ver 2.11.2. This method is based on multivariate logistic regression. We considered: dominant, the heterozygote genotype when it had the same risk effect of the high risk homozygote genotype; recessive, the heterozygote genotype when it had the same risk effect of the low risk homozygote genotype; additive, the heterozygote genotype when it had a risk effect intermediate between the two homozygote genotypes. Firstly, we evaluated the monovariate association between each SNP and the disease. Secondly, among those resulted to be significantly associated with CD, we evaluated the SNP-SNP interactions by the logistic regression model. Finally, the interaction between significant SNPs and HLA groups was verified by an interaction test.

To evaluate the joint effect of HLA and SNPs to influence the risk to develop CD in sibs, we applied a Bayesian approach, focusing only on those SNPs resulted significantly associated with CD at TDT analysis. Usually Bayes' revision of probabilities allows the computation of an individual probability of an event given the *a priori* data from the general population where the patient comes from. We arbitrarily selected an *a priori* probability of 0.5, because in our situation we could not apply the Bayes' revision of probabilities and obtain an a posteriori risk of CD in sibs. Indeed the estimated 10% of CD risk in first grade relatives also account for the role of genetics, therefore it cannot be used as an a priori probability to calculate the a posteriori probability, after considering the risk conferred by genetic factors (SNPs).

Firstly we considered the different frequencies of HLA genotypes in probands and controls and secondly the non-HLA SNPs genotypes

frequencies. According to the Bayesian approach, in each run we used the score obtained in the previous step as the new *a priori* value for the following step. By this approach, it was possible to assign to HLA only or to each combination of HLA plus LPP, RGS1 and REL SNPs genotypes a Bayesian Score (BS). Secondly, we assigned to each sib (testing set) a BS in dependence of the specific HLA-SNPs genotype combination found. We arbitrarily established the median BS as discrimination threshold between low and high risk sibs and evaluated how affected and unaffected sibs were distributed.

## Results

### ***SNPs evaluation and computation of the Bayesian Score***

All individuals were typed for 10 CD associated SNPs (Dubois et al., 2010) and the TDT test was performed on results obtained from 157 trios of the training set (Figure 12). Three out of ten investigated SNPs (rs1464510 in LPP, rs842647 in REL and rs2816316 in RGS1 genes) were significantly associated with CD (Table 8). In particular: rs1464510 in LPP gene showed a strong association ( $p < 0.001$ ) according to an additive model, whereas, both rs2816316 in RGS1 and rs842647 in REL genes were also significantly associated (respectively  $p = 0.025$  and  $p = 0.034$ ), by a recessive model (Table 15).

SNP	Gene	Model	Risk Allele	OR (95% CI)	p value
rs1464510	LPP	Additive	A	2.36 (1.64-3.41)	<0.001
rs2816316	RGS1	Recessive	A	1.75 (1.07-2.86)	0.025
rs842647	REL	Recessive	A	1.66 (1.04-2.65)	0.034
rs2327832	OLIG3	Additive	G	1.35 (0.90-2.03)	0.150
rs6441961	CCR1/CCR3	Additive	A	1.24 (0.89-1.72)	0.189
rs6822844	IL2/IL21	Additive	C	1.43 (0.82-2.49)	0.210
rs1738074	TAGAP	Dominant	A	1.31 (0.79-2.16)	0.293
rs3184504	SH2B3	Additive	A	1.19 (0.86-1.63)	0.294
rs17810546	IL12A	Additive	G	1.10 (0.80-1.51)	0.572
rs9811792	IL12A/SCHIP1	Dominant	G	1.10 (0.59-2.05)	0.753

Table 15. Genotypic Transmission Disequilibrium Test (TDT) results.

The other seven investigated SNPs, even if not statistically associated with CD, showed always an higher frequency of the previously reported risk alleles (Dubois et al., 2010) in affected subjects than in controls.

		Training set	
		Probands	Parents
LPP	A*	0.52	0.43
	C	0.48	0.57
OLIG3	A	0.76	0.79
	G*	0.24	0.21
RGS1	A*	0.88	0.85
	C	0.12	0.15
REL	A*	0.77	0.75
	G	0.23	0.25
SH2B3	A*	0.55	0.53
	G	0.45	0.47
CCR	A*	0.41	0.39
	G	0.59	0.61
IL2/IL21	A	0.09	0.11
	C*	0.91	0.89
IL12A	A	0.57	0.58
	G*	0.43	0.42
TAGAP	A*	0.42	0.41
	G	0.58	0.59
SCHIP1	A	0.92	0.93
	G*	0.08	0.07

Table 16. allelic frequencies observed in Training set (Trios). \*Associated allele.

In order to evaluate the occurrence of HLA-SNP interaction, we stratified the training set (Figure 9) according to the HLA risk group of the

proband. No statistically significant interaction was found between HLA and non-HLA genes (data not shown).

To compute the Bayesian Score (BS) we compared the frequency of each HLA+SNPs genotype combination detected in probands and in controls (Training set, Figure 9). Through this approach we obtained a BS for each HLA+SNPs genotype combination (data not shown).

### ***Validation of the BS and testing of a classification model***

The validation set (Figure 12) was composed by the sibs of the probands, both affected (n=29) and unaffected (n=220). In these subjects we evaluated if HLA+SNPs genotyping could improve the identification of CD risk in sibs better than HLA only.

We assigned to each sib, both affected and unaffected, the BS value corresponding to his HLA+SNPs genotype combination as previously calculated in the training set (Table 17).

Associated SNPs			BS				
LPP (rs1464510)	REL (rs842647)	RGS1 (rs2816316)	HLA Group 1	HLA Group 2	HLA Group 3	HLA Group 4	HLA Group 5
CC	AG GG	AC CC	0.66	0.63	0.32	0.30	0.04
CC	AG GG	AA	0.74	0.73	0.42	0.39	0.06
CC	AA	AC CC	0.71	0.69	0.38	0.35	0.05
CC	AA	AA	0.79	0.77	0.48	0.46	0.08
AC	AG GG	AC CC	0.81	0.80	0.52	0.49	0.09
AC	AG GG	AA	0.87	0.86	0.62	0.60	0.13
AC	AA	AC CC	0.85	0.84	0.58	0.55	0.11
AC	AA	AA	0.90	0.89	0.68	0.66	0.16
AA	AG GG	AC CC	0.91	0.90	*	0.69	0.18
AA	AG GG	AA	0.94	0.93	0.79	0.77	0.25
AA	AA	AC CC	0.93	0.92	0.76	0.74	0.22
AA	AA	AA	0.95	0.95	0.83	0.81	0.30

Table 17. Bayesian Score (BS) assigned to each HLA-SNPs genotype combination.

\*The combination LPP\*AA - REL\*AG|GG - RGS1\*AC|CC for HLA group 3 was not found among our sibs' cohort.

We observed an increase of average BS values from HLA group 5 to HLA group 1 and, within all 5 HLA groups we found an increase of the BS corresponding to an increase of “A” alleles in the haplotype combination. In order to identify sibs at high risk to develop CD, we distributed affected and unaffected sibs on the basis of their BS (under or above median BS) within each HLA group from G1 to G5 (Figure 13).

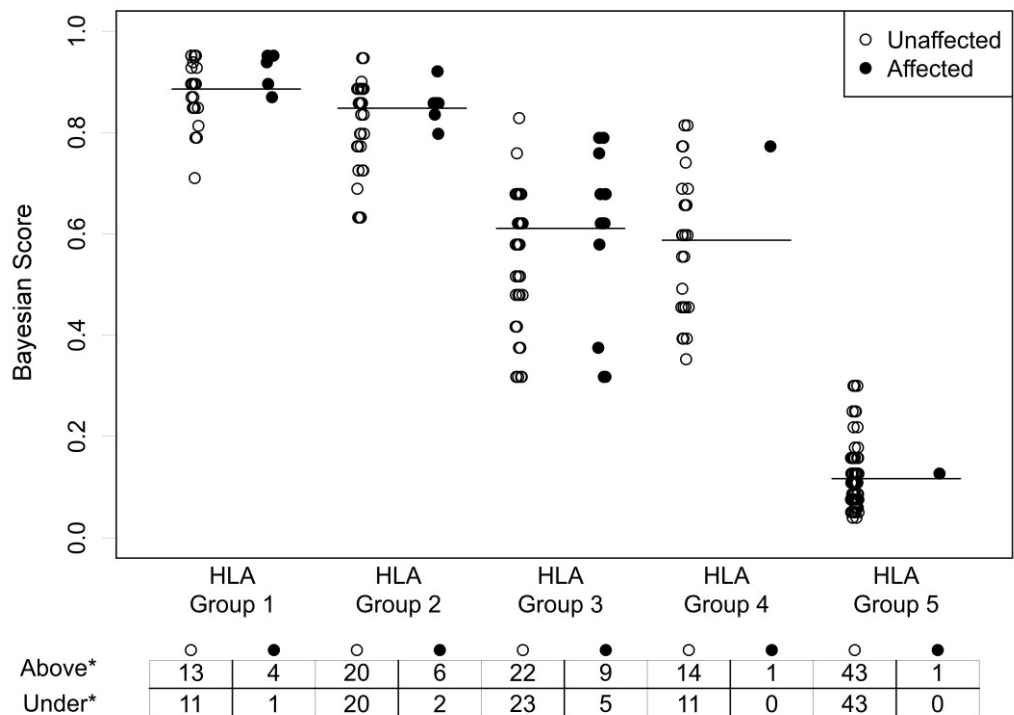


Figure 13. Distribution of CD sibs based on their BS. Affected (n=29) and unaffected (n=220) CD sibs were classified on the basis of their BS value < or ≥ median BS within each HLA group. Horizontal lines correspond to the median BS of all sibs in each HLA risk-group (HLA Group 1 = 0.90, HLA Group 2 = 0.86, HLA Group 3 = 0.62, HLA Group 4 = 0.60, HLA Group 5 = 0.13).

Considering the distribution of all the sibs, it is evident that above the median BS within each HLA group, there was always a larger number of affected sibs than under the median: 72% (21/29) versus 28% (8/29). Interestingly, 2/29 affected sibs that being at lower HLA risk (HLA group 4 and 5) could be misclassified, were correctly classified by their BS above

the median. We calculated the Odds Ratio (OR) of CD classification based on the BS, above or under the median, within each HLA group. CD sibs with a BS value above the median, had more than two fold higher risk (OR) compared to CD sibs with a BS value under the median (2.53, 95% C.I.: 1.68-3.39;  $p=0.047$ ).

In our previous work, we estimated the HLA related risk to develop CD, defining a risk range from 0.01 to 0.21 (*Bourgey et al, 2007*). To refine the previous HLA based estimation of risk by this new approach we set the HLA related mean risk as the a priori risk of the Bayesian Model. Having set the HLA risk at the median level of the Bayesian Score, we produced a picture of the variation of risk given by the non-HLA genes for each HLA class (Figure 14).

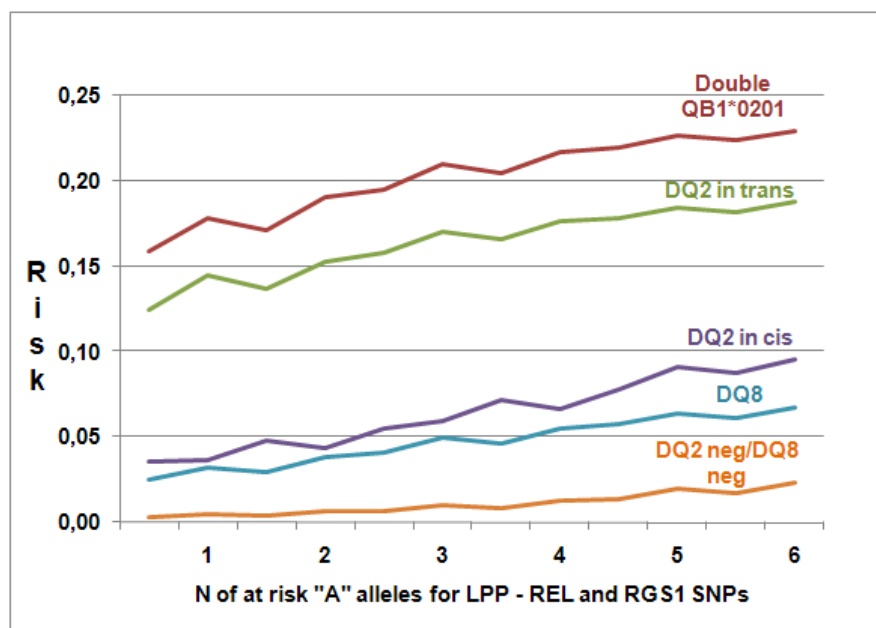


Figure 14. Refining the CD risk estimate. The picture shows the modification of the *a priori* HLA related risk by the number of the at risk "A" alleles of the LPP, REL and RGS1 SNPs. From top to bottom lines correspond to HLA group 1 to 5.

It is remarkable that the addition of the 3 SNPs does modify the HLA only risk through the 5 HLA risk classes, also in subjects with no DQ2

neither DQ8 haplotype, where the increase in risk is significantly larger than that observed in HLA risk classes.

The addition of only 3 SNPs to the HLA significantly improved the prediction of CD risk in sibs, identifying, within a specific HLA group, those individuals which are more likely to become celiacs. In fact, considering HLA 1 and 2 as the highest risk groups (Bourgey et al., 2007), we calculated the diagnostic characteristics of our proposed HLA+SNPs genotype combination obtaining both an higher predictive negative value (NPV) and an higher diagnostic sensitivity (DS) than HLA only, respectively 95% vs 91% (NPV) and 79% vs 45% (DS) (Table 18).

	HLA risk groups 1-2	HLA-SNPs genotype combination
<b>Sensitivity</b>	0.45 (0.27-0.64)	0.79 (0.73-0.84)
<b>Specificity</b>	0.71 (0.65-0.77)	0.54 (0.48-0.60)
<b>NPV</b>	0.91 (0.87-0.96)	0.95 (0.92-0.98)
<b>PPV</b>	0.17 (0.12-0.22)	0.19 (0.14-0.24)

Table 18. diagnostic characteristics of HLA and HLA-SNPs genotype combination.

Although the discovery of 39 polymorphism associated to CD improved the estimation of the heritability of only 4-5% (Dubois et al., 2010), in this Bayesian model the addition of only three SNPs of associated genes improves the sensitivity of risk prediction of 34% compared to the HLA only model.

The computation of the AUC of the ROC curve output a C statistic equal to 0.70 for HLA and 0.73 for HLA+SNPs classifications, showing that the inclusion of SNPs moderately improved the prediction ability.



## Discussion

In our previous work, we considered in CD families the risk to develop the disease according to a specific HLA haplotypes, obtaining a risk range from 0.01 to  $> 0.20$  (Bourgey et al., 2007). In the present study we evaluated the role of 3 non-HLA genetic markers to influence the CD risk in first relatives of CD affected children.

We collected data on families with at least one CD-affected offspring. This family set helped to evaluate the association between SNPs and CD (TDT on parents-offspring trios) and to estimate the risk of CD in the other sibs. The TDT design provides robustness to population stratification and mitigation of the possible confounding effect of environmental factors, because all family members share the same environment (Schaid, 1996).

Ten SNPs, selected from those previously found to be associated with CD by GWAS (Dubois et al., 2010), were successfully genotyped. In our population three SNPs resulted significantly associated with CD (those in LPP, RGS1 and REL genes) and the other seven investigated SNPs, even if not statistically associated with CD, showed always an higher frequency of the previously reported risk alleles in affected subjects than in controls.

The three genes selected appear to be appealing for the pathogenesis of CD. LPP (OR=2.36;  $p<0.001$ ) was reported to be highly expressed in small intestinal mucosa and may have a structural role at sites of cell adhesion in maintaining cell shape and motility (Hunt et al., 2008). RGS1 (OR=1.75;  $p=0.025$ ) belongs to a family of RGS genes. It attenuates the signaling activity of G-proteins, blocking the homing of Intra Epithelial Lymphocytes (IELs), and it is specifically expressed both in human small intestinal mucosa and in murine IELs, key players in the development of human CD villous atrophy (Hunt et al., 2008; Pennington et al., 2003). REL (OR=1.66;  $p=0.034$ ) is a subunit of NF- $\kappa$ B complex, implicated in T cell

differentiation (Son et al., 2011) and it appears to be a key molecule regulating inflammation and the switch from tolerance to autoimmunity (Liou & Smith, 2011). It is interesting to note that our data confirm previous pathogenic implications reported in the literature of these SNPs with CD as well as with other autoimmune diseases (Abadie et al., 2011).

By the Bayesian approach we calculated a ranking score (BS) among the sibs. However, it should be considered that BS is not a plain disease risk, rather a method to rank different genotypes according to their contribution to individual susceptibility to CD. For instance, some of our BS are very near to 1, nevertheless none of the considered genotypes could give a 100% risk to develop the disease. In other terms, we considered the BS as a ranking measure, only stating that a given genotype could assign a higher risk than another genotype but does not allow a quantitative measure of the risk difference (2-fold, 3-fold, etc). However, even if the addition of only 3 SNPs to HLA could be considered at “minor effect”, we demonstrated that they could significantly improve the prediction of CD risk in sibs, in terms of diagnostic sensitivity and negative predictive value. So, in a cohort of CD families, our data confirm that non-HLA SNPs evaluation is an useful diagnostic tool in CD risk evaluation as a previous study showed in CD unrelated subjects (Romanos et al., 2009).

CD, on the basis of the actual knowledge, cannot be exactly predicted by genetic testing, but a reliable probabilistic method might be associated to careful surveillance of infants carrying the higher risk. This will help to significantly reduce the heavy load of anxiety and pain associated with the appearance of symptoms of CD, by anticipating, with simple serological tests, the clinical appearance of the disease.

To improve the possibility to identify high risk patients in CD families we propose in alternative to the classical HLA classification (Figure 11) a slight improved flow-chart (Figure 12).

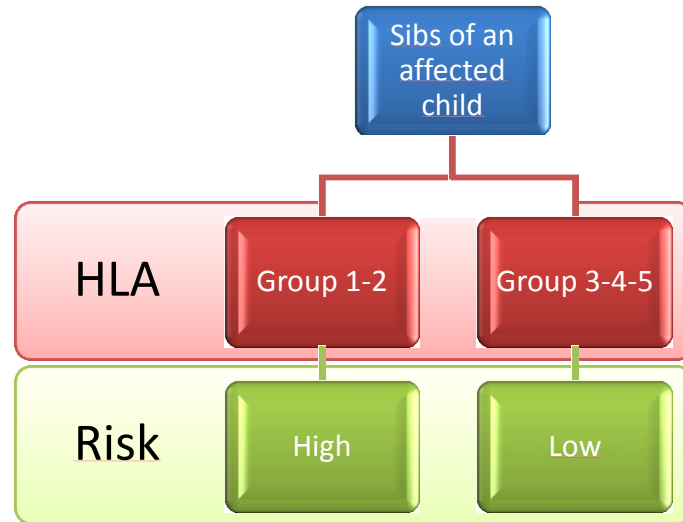


Figure 11. Classification flow-chart. The classical HLA-based classification.

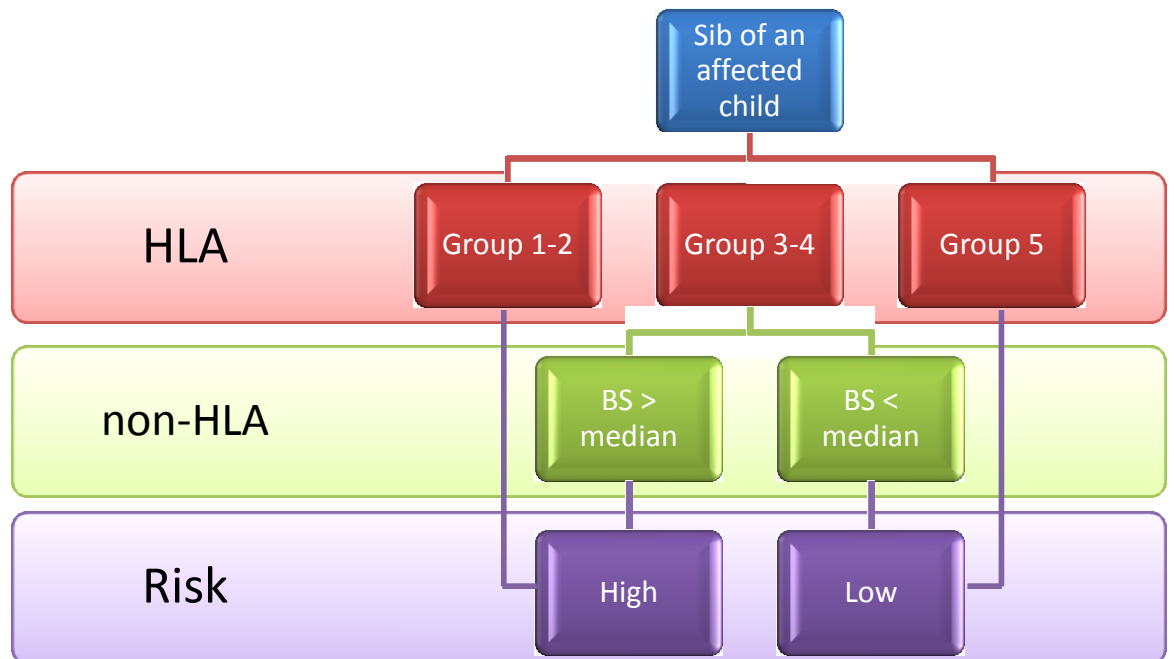


Figure 12. The proposed BS-based classification considering the genotypes of HLA plus LPP, RGS1 and REL SNPs.

- 1) HLA genotyping: subjects belonging to the HLA risk groups 1 and 2 will be classified as at high CD risk;
- 2) subjects belonging to the HLA risk groups 3 and 4, will be further investigated for our SNPs combination (LPP, REL, RGS1) in order to calculate their BS (Figure 12). Among these latter subjects those with a BS > the median value will be classified at high risk;
- 3) subjects belonging to the HLA risk group 5 will be considered at low CD risk. All CD familiars belonging to the above high risk groups (HLA group 1-2 and HLA group 3-4 with BS > median) will be undergo a strict surveillance.

One of the limitation of our family study could be the sample size, which may have not allowed to explore genes at smaller effect, so explaining the lack of association between SNPs in TAGAP, IL2/IL21, OLIG3, CCR, SH2B3, IL12A and IL12A/SCHIP1 genes with CD although the trend observed in previous studies in unrelated CD patients was confirmed (Dubois et al., 2010). In the main time the homogeneity of the genetic and environmental domains in the tested families allows to explore risk factors within a controlled cohort. A second limit of the study is the relatively short (6 years) follow up of the sibship, which could cause an underestimation of the disease development at later ages. Our aim is to go on with the monitoring of these families in the next years.

In conclusion, the estimate of the CD risk by HLA+SNPs approach, even if not applicable to prevention, could be a precious tool improving CD diagnosis respect to the only HLA (NPV: 95% vs 91%, and DS: 79% vs 45%), in the cohort of first degree relatives. In fact in clinical practice the absence of HLA risk groups 1 or 2, allows to exclude the disease with high probability, while testing the three SNPs in HLA groups 3 or 4 could represent a further tool to identify less frequent CD cases. So, an infant with high HLA+SNPs score even if belonging to HLA low risk groups, shall

undergo a simple surveillance system to allow proper diagnosis and treatment before the full blown disease appearing.

**General**

**Conclusions and Perspectives**

## **General conclusions and perspective**

Our aim is to extract as many information as we can from the large amount of data emerged from GWAs. Even though some research groups are investing on this issue, the discovery of new associated genes does not prove any pathogenic mechanism behind the development of CD. In fact the rationale for GWAS is the 'common disease, common variant', where common variant are those present in more than 5% of the population (Manolio et al., 2009). For this reason, SNP chips or array cannot evaluate deletion and duplication (carried by 8% of the general population) and copy-number variation (CNV) inaccessible by most existing genotyping and sequencing technologies. Furthermore Gene-Environment (GxE) or Gene-Gene (GxG) interactions (Hindorff et al., 2009; Manolio et al., 2009), not to mention the possible epigenetic and regulation (microRNA) effects (Eichler et al., 2010; Slatkin, 2009), cannot be revealed by GWAS and despite their ability to detect the role of genetic factors in complex diseases, literature about this issue is still poor. Next generation sequencing technologies could be performed in extreme phenotypes in order to identify other associated variants, both common and rare. But again, the accumulation of data from GWAs and from others investigation methods will lead, on the long run, to consider genes associated to CD genes encoding proteins involved in immunity and inflammation, describing the physiological cell functions, without bringing new substantial data to rewrite the pathogenesis of the disease.

It would be more appropriate to focus only on some of them, to be sure that "making a list of genes" wasn't seen in an un-useful way. During my PhD training, my group and I tried to speculate in a critical way around the non-HLA genes issue. For this reason my PhD thesis project is articulated at various points.

Firstly, we tried to improve the analyses of candidate genes analyzing their expression profile and comparing the results obtained on mucosal biopsies, peripheral monocytes and T-cell lines isolated from duodenal mucosa: three different cell population with three different functions and locations, in order to give a picture of the variation in the celiac type and to correlate it with a model of CD pathogenesis. In this way we were able to localize the role of these genes in one area rather than another. After that, stimulated by the recently proposed discriminant analysis of celiac gene expression as a promising diagnostic tool with which to distinguish celiac mucosa from normal mucosa (Bragde et al.,2011), we evaluated the diagnostic strength of this analysis on the gene expression data obtained from peripheral blood monocytes. Our final aim was to provide a less invasive alternative to duodenal biopsy as the only diagnostic tool for CD. Using a discriminant analysis, we demonstrate that the expression of a small set of candidate genes in peripheral blood monocytes can be used to correctly classify CD patients from controls and from patients affected by Crohn disease, without considering HLA and anti-tTG levels, because otherwise the diagnostic power of our method could have been hidden by the very high sensitivity and specificity of the former and the very high negative predictive value of the latter.

Speaking in an over-optimistically way, this should be done in clinical practice in order to increase the sensitivity and the specificity of the diagnosis when duodenal biopsy is not available.

Secondly, we aimed to investigate the non-HLA genetics of Potential CD patients. The term '*potential CD*' was assigned to individuals with proper DQ2 or DQ8, production of anti-tTG, normal small intestinal mucosa, classified as Marsh 0 (no damage) or Marsh 1 stage (unspecific intra epithelial infiltration only). These are also defined "serological positive cases" (O. Koskinen et al., 2008; Kurppa et al., 2010). Potential patients



suggest that the development of adaptive anti-gluten immunity is not sufficient to develop villous atrophy. This is supported by animal models showing that inflammatory anti-gluten responses are not sufficient to develop mucosal lesions. Thus, we can subgroup individuals into three groups: controls, patients who developed antibodies but no intestinal lesions (M0 and M1 potentials) and patients with antibodies and villous atrophy (overt CD). Analyzing differences among these subgroups may shed light into the genes differentially involved in development of adaptive anti-gluten immunity (different between control and potential) and tissue lesions (different between potential and CD), as well as genes involved in all steps of CD pathogenesis (different between control and potential, and potential and CD).

Potential patients have most of the features of CD patients: they have similar genes, produce the same antibodies but they lack the final destructive phase of the gluten induced immunoresponse, the cytotoxic destruction of small intestinal mucosa. No clinical, serological or histological features could distinguish M0 from M1 patients up to now (Tosco et al., 2011). For the first time we can speculate that potential CD are a mixture of two populations, with different non-HLA genetics and significant differential expression of some candidate genes.

Thirdly, we tried to use the new genetic findings in order to build a risk model for CD using HLA and non-HLA risk alleles and to significantly improve the identification of subjects at risk to develop CD. We developed a classification system based on a score (Bayesian Score) which can be easily calculated on the basis of three family-associated genes (LPP, c-REL and RGS1) and that permits to discriminate between patients at higher or lower risk to develop the disease. In this way, families with one affected

case will know, with reliable accuracy, the individual risk of a newborn and will be able to anticipate the onset of clinical phenotype by a simple point-of-care tests.

In conclusion, the critic analysis of the large panorama of CD candidate genes leads to consider several mechanisms of action that may have gone unnoticed if we persist in considering all the genes together. Whereas, analyzing a small set of genes from different points of view it is possible to really understand why this or that gene has resulted to be associated with the disease.

GWAs made the history of complex diseases, but represent only the starting point from which postulate hypothesis, mechanisms of action, interactions, in order to untangle the skein represented by the complex pathogenesis of celiac disease.

## References

- Abadie, V., Sollid, L. M., Barreiro, L. B., & Jabri, B. (2011). Integration of genetic and immunological insights into a model of celiac disease pathogenesis. *Annual review of immunology*, 29, 493–525.
- Agostoni, C., Decsi, T., Fewtrell, M., Goulet, O., Kolacek, S., Koletzko, B., Michaelsen, K. F., et al. (2008). Complementary feeding: a commentary by the ESPGHAN Committee on Nutrition. *Journal of pediatric gastroenterology and nutrition*, 46, 99–110.
- Amundsen, S. S., Rundberg, J., Adamovic, S., Gudjónsdóttir, a H., Ascher, H., Ek, J., Nilsson, S., et al. (2010). Four novel coeliac disease regions replicated in an association study of a Swedish-Norwegian family cohort. *Genes and immunity*, 11, 79–86.
- Bevan, S., Popat, S., Braegger, C. P., Busch, A., O'Donoghue, D., Falth-Magnusson, K., Ferguson, A., et al. (1999). Contribution of the MHC region to the familial risk of coeliac disease. *Journal of medical genetics*, 36(9), 687–690.
- Bonamico, M., Ferri, M., Mariani, P., Nenna, R., Thanasi, E., Luparia, R. P. L., Picarelli, A., et al. (2006). Serologic and genetic markers of celiac disease: a sequential study in the screening of first degree relatives. *Journal of pediatric gastroenterology and nutrition*, 42, 150–154.
- Bonizzi, G., & Karin, M. (2004). The two NF-kappaB activation pathways and their role in innate and adaptive immunity. *Trends in immunology*, 25, 280–288.
- Bourgey, M., Calcagno, G., Tinto, N., Gennarelli, D., Margaritte-Jeannin, P., Greco, L., Limongelli, M. G., et al. (2007). HLA related genetic risk for coeliac disease. *Gut*, 56, 1054–1059.
- Bragde, H., Jansson, U., Jarlsfelt, I., & Söderman, J. (2011). Gene expression profiling of duodenal biopsies discriminates celiac disease mucosa from normal mucosa. *Pediatric research*, 69, 530–537.
- Chen, R., Stahl, E. A., Kurreeman, F. A. S., Gregersen, P. K., Siminovitch, K. A., Worthington, J., Padyukov, L., et al. (2011). Fine mapping the TAGAP risk locus in rheumatoid arthritis. *Genes and immunity*, 12(4), 314–318.
- Cohavy, O., Zhou, J., Granger, S. W., Ware, C. F., & Targan, S. R. (2004). LIGHT expression by mucosal T cells may regulate IFN-gamma

expression in the intestine. *Journal of immunology (Baltimore, Md. : 1950)*, 173, 251–258.

De Nitto, D., Monteleone, I., Franzè, E., Pallone, F., & Monteleone, G. (2009). Involvement of interleukin-15 and interleukin-21, two gamma-chain-related cytokines, in celiac disease. *World journal of gastroenterology : WJG*, 15, 4609–4614.

Dema, B., Martínez, A., Fernández-Arquero, M., Maluenda, C., Polanco, I., de la Concha, E. G., Urcelay, E., et al. (2009). Association of IL18RAP and CCR3 with coeliac disease in the Spanish population. *Journal of medical genetics*, 46, 617–619.

Dubois, P. C. A., Trynka, G., Franke, L., Hunt, K. A., Romanos, J., Curtotti, A., Zhernakova, A., et al. (2010). Multiple common variants for celiac disease influencing immune gene expression. *Nature genetics*, 42, 295–302.

Dubé, C., Rostom, A., Sy, R., Cranney, A., Saloojee, N., Garritty, C., Sampson, M., et al. (2005). The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology*, 128, S57–67.

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature reviews. Genetics*, 11, 446–450.

Fina, D., Sarra, M., Caruso, R., Del Vecchio Blanco, G., Pallone, F., MacDonald, T. T., & Monteleone, G. (2008). Interleukin 21 contributes to the mucosal T helper cell type 1 response in coeliac disease. *Gut*, 57, 887–892.

Gianfrani C, Siciliano R, Facchiano A, Camarca A, Mazzeo M, Costantini S, Salvati V, Maurano F, Mozzarella G, Iaquinto G, Bergamo P, R. M. (2007). Transamidation of wheat flour inhibits the response to gliadin of intestinal T cells in celiac disease. *Gastroenterology*, 133, 780–789.

Greco, L., Romino, R., Coto, I., Di Cosmo, N., Percopo, S., Maglio, M., Paparo, F., et al. (2002). The first large population based twin study of coeliac disease. *Gut*, 50, 624–628.

Gueller, S., Hehn, S., Nowak, V., Gery, S., Serve, H., Brandts, C. H., & Koeffler, H. P. (2011). Adaptor protein Lnk binds to PDGF receptor and

inhibits PDGF-dependent signaling. *Experimental hematology*, 39, 591–600.

Han, S.-B., Moratz, C., Huang, N.-N., Kelsall, B., Cho, H., Shi, C.-S., Schwartz, O., et al. (2005). Rgs1 and Gnai2 regulate the entrance of B lymphocytes into lymph nodes and B cell motility within lymph node follicles. *Immunity*, 22, 343–354.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 9362–9367.

Hunt, K. A., Zhernakova, A., Turner, G., Heap, G. A. R., Franke, L., Bruinenberg, M., Romanos, J., et al. (2008). Newly identified genetic risk variants for celiac disease related to the immune response. *Nature genetics*, 40, 395–402.

Husby, S., Koletzko, S., Korponay-Szabó, I. R., Mearin, M. L., Phillips, A., Shamir, R., Troncone, R., et al. (2012). European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *Journal of pediatric gastroenterology and nutrition*, 54, 136–160.

Izzo, V., Pinelli, M., Tinto, N., Esposito, M. V., Cola, A., Sperandeo, M. P., Tucci, F., et al. (2011). Improving the estimation of celiac disease sibling risk by non-HLA genes. *PloS one*, 6, e26920.

Kaukinen, K., Partanen, J., Mäki, M., & Collin, P. (2002). HLA-DQ typing in the diagnosis of celiac disease. *The American journal of gastroenterology*, 97, 695–699.

Kolodziej, L. E., Lodolce, J. P., Chang, J. E., Schneider, J. R., Grimm, W. A., Bartulis, S. J., Zhu, X., et al. (2011). TNFAIP3 maintains intestinal barrier function and supports epithelial cell tight junctions. *PloS one*, 6, e26352.

Kondrashova, A., Mustalahti, K., Kaukinen, K., Viskari, H., Volodicheva, V., Haapala, A.-M., Ilonen, J., et al. (2008). Lower economic status and inferior hygienic environment may protect against celiac disease. *Annals of medicine*, 40, 223–231.

Koskinen, L. L. E., Einarsdottir, E., Dukes, E., Heap, G. A. R., Dubois, P., Korponay-Szabo, I. R., Kaukinen, K., et al. (2009). Association study of

the IL18RAP locus in three European populations with coeliac disease. *Human molecular genetics*, 18, 1148–1155.

Koskinen, O., Collin, P., Korponay-Szabo, I., Salmi, T., Iltanen, S., Haimila, K., Partanen, J., et al. (2008). Gluten-dependent small bowel mucosal transglutaminase 2-specific IgA deposits in overt and mild enteropathy coeliac disease. *Journal of pediatric gastroenterology and nutrition*, 47, 436–442.

Kuo, M. T., Wei, Y., Yang, X., Tatebe, S., Liu, J., Troncoso, P., Sahin, A., et al. (2006). Association of fragile site-associated (FSA) gene expression with epithelial differentiation and tumor development. *Biochemical and biophysical research communications*, 340, 887–893.

Kurppa, K., Ashorn, M., Iltanen, S., Koskinen, L. L. E., Saavalainen, P., Koskinen, O., Mäki, M., et al. (2010). Celiac disease without villous atrophy in children: a prospective study. *The Journal of pediatrics*, 157, 373–380, 380.e1.

Lawrence, T., Bebie, M., Liu, G. Y., Nizet, V., & Karin, M. (2005). IKK $\alpha$  limits macrophage NF- $\kappa$ B activation and contributes to the resolution of inflammation. *Nature*, 434, 1138–1143.

Liou, H.-C., & Smith, K. A. (2011). The roles of c-rel and interleukin-2 in tolerance: a molecular explanation of self-nonsel discrimination. *Immunology and cell biology*, 89, 27–32.

Liu, Y., Helms, C., Liao, W., Zaba, L. C., Duan, S., Gardner, J., Wise, C., et al. (2008). A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS genetics*, 4, e1000041.

Maiuri, M. C., De Stefano, D., Mele, G., Fecarotta, S., Greco, L., Troncone, R., & Carnuccio, R. (2003). Nuclear factor kappa B is activated in small intestinal mucosa of celiac patients. *Journal of molecular medicine (Berlin, Germany)*, 81, 373–379.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., et al. (2009). Finding the missing heritability of complex diseases. *Nature*, 461, 747–753.

Margaritte-Jeannin, P., Babron, M. C., Bourgey, M., Louka, A. S., Clot, F., Percopo, S., Coto, I., et al. (2004). HLA-DQ relative risks for coeliac disease in European populations: a study of the European Genetics Cluster on Coeliac Disease. *Tissue antigens*, 63, 562–567.

- Mustalahti, K., Catassi, C., Reunanen, A., Fabiani, E., Heier, M., McMillan, S., Murray, L., et al. (2010). The prevalence of celiac disease in Europe: results of a centralized, international mass screening project. *Annals of medicine*, 42, 587–595.
- Oberhuber, G., Caspary, W. F., Kirchner, T., Borchard, F., & Stolte, M. (2001). [Study Group of Gastroenterological Pathology of the German Society of Pathology. Recommendations for celiac disease/sprue diagnosis]. *Zeitschrift für Gastroenterologie*, 39, 157–166.
- Pennington, D. J., Silva-Santos, B., Shires, J., Theodoridis, E., Pollitt, C., Wise, E. L., Tigelaar, R. E., et al. (2003). The inter-relatedness and interdependence of mouse T cell receptor gammadelta+ and alphabeta+ cells. *Nature immunology*, 4, 991–998
- Petit, M. M., Fradelizi, J., Golsteyn, R. M., Ayoubi, T. A., Menichi, B., Louvard, D., Van de Ven, W. J., et al. (2000). LPP, an actin cytoskeleton protein related to zyxin, harbors a nuclear export signal and transcriptional activation capacity. *Molecular biology of the cell*, 11, 117–129.
- Petit, M. M., Mols, R., Schoenmakers, E. F., Mandahl, N., & Van de Ven, W. J. (1996). LPP, the preferred fusion partner gene of HMGIC in lipomas, is a novel member of the LIM protein gene family. *Genomics*, 36, 118–129.
- Petit, M. M. R., Meulemans, S. M. P., & Van de Ven, W. J. M. (2003). The focal adhesion and nuclear targeting capacity of the LIM-containing lipoma-preferred partner (LPP) protein. *The Journal of biological chemistry*, 278, 2157–2168.
- Petronzelli, F., Bonamico, M., Ferrante, P., Grillo, R., Mora, B., Mariani, P., Apollonio, I., et al. (1997). Genetic contribution of the HLA region to the familial clustering of coeliac disease. *Annals of human genetics*, 61, 307–317.
- Plaza-Izurieta, L., Castellanos-Rubio, A., Irastorza, I., Fernández-Jimenez, N., Gutierrez, G., & Bilbao, J. R. (2011). Revisiting genome wide association studies (GWAS) in coeliac disease: replication study in Spanish population and expression analysis of candidate genes. *Journal of medical genetics*, 48, 493–496.
- Plot, L., Amital, H., Barzilai, O., Ram, M., Nicola, B., & Shoenfeld, Y. (2009). Infections may have a protective role in the etiopathogenesis of celiac disease. *Annals of the New York Academy of Sciences*, 1173, 670–674.

- Romanos, J, Barisani, D., Trynka, G., Zhernakova, A., Bardella, M. T., & Wijmenga, C. (2009). Six new coeliac disease loci replicated in an Italian population confirm association with coeliac disease. *Journal of medical genetics*, 46, 60–63.
- Romanos, Jihane, van Diemen, C. C., Nolte, I. M., Trynka, G., Zhernakova, A., Fu, J., Bardella, M. T., et al. (2009). Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. *Gastroenterology*, 137, 834–840, 840.
- Rostami, K., Kerckhaert, J., Tiemessen, R., von Blomberg, B. M., Meijer, J. W., & Mulder, C. J. (1999). Sensitivity of antiendomysium and antigliadin antibodies in untreated celiac disease: disappointing in clinical practice. *The American journal of gastroenterology*, 94, 888–894.
- Schaid, D. J. (1996). General score tests for associations of genetic markers with disease using cases and their parents. *Genetic epidemiology*, 13, 423–449.
- Shaikh, R. B., Santee, S., Granger, S. W., Butrovich, K., Cheung, T., Kronenberg, M., Cheroutre, H., et al. (2001). Constitutive expression of LIGHT on T cells leads to lymphocyte activation, inflammation, and tissue destruction. *Journal of immunology (Baltimore, Md. : 1950)*, 167, 6330–6337.
- Shi, G.-X., Harrison, K., Han, S.-B., Moratz, C., & Kehrl, J. H. (2004). Toll-like receptor signaling alters the expression of regulator of G protein signaling proteins in dendritic cells: implications for G protein-coupled receptor signaling. *Journal of immunology (Baltimore, Md. : 1950)*, 172, 5175–5184.
- Slatkin, M. (2009). Epigenetic inheritance and the missing heritability problem. *Genetics*, 182, 845–850
- Sollid, L. M., Markussen, G., Ek, J., Gjerde, H., Vartdal, F., & Thorsby, E. (1989). Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *The Journal of experimental medicine*, 169, 345–350.
- Son, J.-S., Sahoo, A., Chae, C.-S., Hwang, J.-S., Park, Z. Y., & Im, S.-H. (2011). JunB and c-Rel cooperatively enhance Foxp3 expression during induced regulatory T cell differentiation. *Biochemical and biophysical research communications*, 407, 141–147.



- Sperandeo, M. P., Tosco, A., Izzo, V., Tucci, F., Troncone, R., Auricchio, R., Romanos, J., et al. (2011). Potential celiac patients: a model of celiac disease pathogenesis. *PloS one*, 6, e21281
- Stahl, E. A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B. F., Kraft, P., et al. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genetics*, 44, 483–489.
- Takaki, S. (2008). [Sh2b3/Lnk family adaptor proteins in the regulation of lymphohematopoiesis]. *Nihon Rinshō Men'eki Gakkai kaishi = Japanese journal of clinical immunology*, 31, 440–447.
- Thomson, G. (1995). Mapping disease genes: family-based association studies. *American journal of human genetics*, 57, 487–498.
- Todd, J. A., Walker, N. M., Cooper, J. D., Smyth, D. J., Downes, K., Plagnol, V., Bailey, R., et al. (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature genetics*, 39, 857–864.
- Tosco, A., Salvati, V. M., Auricchio, R., Maglio, M., Borrelli, M., Coruzzo, A., Paparo, F., et al. (2011). Natural history of potential celiac disease in children. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*, 9, 320–325.
- Trynka, G., Zhernakova, A., Romanos, J., Franke, L., Hunt, K. A., Turner, G., Bruinenberg, M., et al. (2009). Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. *Gut*, 58,
- Wei, Y., Lin-Lee, Y.-C., Yang, X., Dai, W., Zhao, S., Rassool, F. V., Elgart, G. W., et al. (2006). Molecular cloning of Chinese hamster 1q31 chromosomal fragile site DNA that is important to *mdr1* gene amplification reveals a novel gene whose expression is associated with spermatocyte and adipocyte differentiation. *Gene*, 372, 44–52.
- Wertz, I. E., O'Rourke, K. M., Zhou, H., Eby, M., Aravind, L., Seshagiri, S., Wu, P., et al. (2004). De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-kappaB signalling. *Nature*, 430, 694–699.
- Williams, M. E., Chang, T. L., Burke, S. K., Lichtman, A. H., & Abbas, A. K. (1991). Activation of functionally distinct subsets of CD4+ T lymphocytes. *Research in immunology*, 142, 23–28.

- Working Group of European Society of Paediatric Gastroenterology and Nutrition. (1990). Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Archives of disease in childhood*, 6, 909–911.
- Yamanouchi, J., Rainbow, D., Serra, P., Howlett, S., Hunter, K., Garner, V. E. S., Gonzalez-Munoz, A., et al. (2007). Interleukin-2 gene variation impairs regulatory T cell function and causes autoimmunity. *Nature genetics*, 39, 329–337.
- Zhernakova, A., Alizadeh, B. Z., Bevova, M., van Leeuwen, M. A., Coenen, M. J. H., Franke, B., Franke, L., et al. (2007). Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. *American journal of human genetics*, 81, 1284–1288.
- van Heel, D. A., Franke, L., Hunt, K. A., Gwilliam, R., Zhernakova, A., Inouye, M., Wapenaar, M. C., et al. (2007). A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nature genetics*, 39, 827–829.
- van de Wal, Y., Kooy, Y., van Veelen, P., Peña, S., Mearin, L., Papadopoulos, G., & Koning, F. (1998). Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. *Journal of immunology (Baltimore, Md. : 1950)*, 161, 1585–1588.

## **Appendix I – Curriculum Vitae**

## Curriculum Vitae

### Personal information

First name / Surname **Valentina Izzo**  
Address Via Aldo Moro 21, 80018, Mugnano di Napoli, Napoli, Italy  
Telephone +39 081 745 22 85 Mobile 1: +39 333 46 22 805  
Mobile 2: +33 6 50 16 29 07  
Fax  
E-mail [valent.izzo@gmail.com](mailto:valent.izzo@gmail.com)  
Nationality Italian  
Date of birth February 25<sup>th</sup> 1986  
Gender Female

### Education and training (1/4)

Dates From March 1<sup>st</sup> 2010 to February 28<sup>th</sup> 2013  
Title of qualification awarded Doctor of Philosophy  
Principal subjects/occupational skills covered PhD student at the Doctoral Course in "Sviluppo, accrescimento e riproduzione dell'uomo" with a research project focused on the genetics of Celiac Disease.  
Name and type of organisation providing education and training Dipartimento di Scienze Mediche Translazionali  
Sezione di Pediatria  
Università di Napoli Federico II  
Referee Prof. Luigi Greco  
Level in national or international classification ISCED 6

### Education and training (2/4)

Dates From April 16<sup>th</sup> 2012 to February 28<sup>th</sup> 2013  
Title of qualification awarded Fellowship  
Principal subjects/occupational skills covered To investigate the possible role of autophagy in Celiac Disease; improving the knowledge about main laboratory expertise.  
Name and type of organisation providing education and training INSERM U848, Institut G. Roussy (PR1) 39 rue C. Desmoulins, Villejuif, Paris, France  
Referee Prof. Guido Kroemer

### Education and training (3/4)

Dates	From December 2007 to October 2009
Title of qualification awarded	Master degree in "Biotechnology applied to the Pharmacology" (II level degree), Italian grade of 110/110 cum laude
Principal subjects/occupational skills covered	Internship focused on the expression levels of Celiac Disease candidate genes in different tissue and cell type
Name and type of organisation providing education and training	Department of Paediatrics, "Federico II" University, Via Pansini 5, 80131, Naples, Italy
Referee	Prof. Luigi Greco
Level in national or international classification	ISCED 5

### Education and training (4/4)

Dates	From September 2004 to December 2007
Title of qualification awarded	Bachelor degree in "Biotechnology for Healthcare (I level degree), Italian grade of 104/110
Principal subjects/occupational skills covered	Internship focused on the role of the GTPase Rac in the cytoskeletal organization of thyroid cancer cells.
Name and type of organisation providing education and training	Department of Molecular and Cell Biology and Pathology "L. Califano", "Federico II" University,
Referee	Prof. Lucio Nitsch
Level in national or international classification	ISCED 5

### Personal skills and competences

Mother tongue **Italian**

Other language(s)

Self-assessment

European level (\*)

**English**

**French**

Understanding		Speaking		Writing	
Listening	Reading	Spoken interaction	Spoken production		
B2	Independent user	B2	Independent user	B2	Independent user
B1	Independent user	B1	Independent user	A1	Basic user

(\*) [Common European Framework of Reference for Languages](#)

Organisational skills and competences	<p>Sense of organization</p> <p>Capacity to design a scientific project.</p> <p>Capacity to administrate small budget for the daily work in a small laboratory</p> <p>Capacity to write a scientific article, a book chapter and a review</p>
Technical skills and competences	<p>Cell culture methods; Cell transfection (stable, transient); Cell Immunofluorescence and detection by confocal microscopy; protein extraction; Western blot; Immunoprecipitation; Dot Blot; DNA extraction and detection from all biological samples; cytofluorimetric analyses; Agarose gel electrophoresis; PCR; RNA extraction and detection from tissue and cells; RT-PCR; Real Time-PCR and analysis by 7300 and 7900 Fast (AB); Allelic Discrimination and analysis by 7900 Fast (AB). Intermediate knowledge of the main statistical analysis.</p>
Computer skills and competences	<p>Knowledge of Entrez gene, genecards, embl nucleotide sequence database, SNP database.</p> <p>Good command of Microsoft Office™ tools (Word™, Excel™ and PowerPoint™)</p> <p>Use of SPSS 20.0 and R (MS-DOS) for data and statistical analysis.</p> <p>Use of GraphPad Prism 6.0 for graph and statistical analysis.</p> <p>Use of Photoshop for image elaboration</p> <p>Use of Mendeley 1.3.1 and EndNote to elaborate reference list in papers and thesis.</p>
References	<p>Prof. Luigi Greco, Dept. of Pediatrics, University of Naples "Federico II", Via Pansini n.5, 80131 Naples, Italy Phone: +39- 81-7463275, e-mail ydongre@unina.it</p> <p>Dr. Maria Chiara Maiuri INSERM U848, Institut G. Roussy (PR1) 39 rue C. Desmoulins, Villejuif, Paris, France e-mail: maiuri@igr.fr</p>
Publications	<p>Capuano M, Iaffaldano L, Tinto N, Montanaro D, Capobianco V, <u>Izzo V</u>, Tucci F, Troncone G, Greco L, Sacchetti L. MicroRNA-449a Overexpression, Reduced NOTCH1 Signals and Scarce Goblet Cells Characterize the Small Intestine of Celiac Patients. PLoS One. 2011;6:e29094.</p> <p><u>Izzo V</u>, Pinelli M, Tinto N, Esposito MV, Cola A, Sperandeo MP, Tucci F, Cocozza S, Greco L, Sacchetti L. Improving the estimation of celiac disease sibling risk by non-HLA genes. PLoS One. 2011;6:e26920.</p> <p>Trynka G, Hunt KA, Bockett NA, et al. <u>Izzo V</u>, et al., Wijmenga C, van Heel DA. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 2011 Nov 6;43:1193-1201.</p> <p>Sperandeo MP, Tosco A, <u>Izzo V</u>, Tucci F, Troncone R, Auricchio R, Romanos J, Trynka G, Auricchio S, Jabri B, Greco L. Potential celiac patients: a model of celiac disease pathogenesis. PLoS One. 2011;6:e21281.</p>

## **Appendix II – Publications**

# Potential Celiac Patients: A Model of Celiac Disease Pathogenesis

Maria Pia Sperandio<sup>1</sup>, Antonella Tosco<sup>1</sup>, Valentina Izzo<sup>1</sup>, Francesca Tucci<sup>1</sup>, Riccardo Troncone<sup>1</sup>, Renata Auricchio<sup>1</sup>, Jihane Romanos<sup>2</sup>, Gosia Trynka<sup>2</sup>, Salvatore Auricchio<sup>1</sup>, Bana Jabri<sup>3</sup>, Luigi Greco<sup>4</sup>\*

<sup>1</sup> European Laboratory for Food Induced Disease, University of Naples Federico II, Naples, Italy, <sup>2</sup> Department of Genetics, University Medical Centre of Groningen, University of Groningen, Groningen, The Netherlands, <sup>3</sup> Department of Medicine, Department of Pathology and Department of Pediatrics, University of Chicago, Chicago, Illinois, United States of America, <sup>4</sup> Department of Pediatrics, University of Naples Federico II, Naples, Italy

## Abstract

**Background and Aim:** Potential celiacs have the 'celiac type' HLA, positive anti-transglutaminase antibodies but no damage at small intestinal mucosa. Only a minority of them develops mucosal lesion. More than 40 genes were associated to Celiac Disease (CD) but we still do not know how those pathways transform a genetically predisposed individual into an affected person. The aim of the study is to explore the genetic features of Potential CD individuals.

**Methods:** 127 'potential' CD patients entered the study because of positive anti-tissue transglutaminase and no mucosal lesions; about 30% of those followed for four years become frankly celiac. They were genotyped for 13 polymorphisms of 'candidate genes' and compared to controls and celiacs. Moreover, 60 biopsy specimens were used for expression studies.

**Results:** Potential CD bear a lighter HLA-related risk, compared to celiac ( $\chi^2 = 48.42$ ;  $p$  value =  $1 \times 10^{-8}$ ). They share most of the polymorphisms of the celiacs, but the frequency of c-REL\* G allele was suggestive for a difference compared to celiac ( $\chi^2 = 5.42$ ;  $p$  value = 0.02). One marker of the KIAA1109/IL-2/IL-21 candidate region differentiated potentials from celiac (rs4374642:  $\chi^2 = 7.17$ ,  $p$  value = 0.01). The expression of IL-21 was completely suppressed in potentials compared to celiacs ( $p$  value = 0.02) and to controls ( $p$  value = 0.02), in contrast IL-2, KIAA1109 and c-REL expression were over-expressed.

**Conclusions:** Potential CD show genetic features slightly different from celiacs. Genetic and expression markers help to differentiate this condition. Potential CD is a precious biological model of the pathways leading to the small intestinal mucosal damage in genetically predisposed individuals.

**Citation:** Sperandio MP, Tosco A, Izzo V, Tucci F, Troncone R, et al. (2011) Potential Celiac Patients: A Model of Celiac Disease Pathogenesis. PLOS ONE 6(7): e21281. doi:10.1371/journal.pone.0021281

**Editor:** Christian Schönbach, Kyushu Institute of Technology, JAPAN

**Received:** March 23, 2011; **Accepted:** May 24, 2011; **Published:** July 8, 2011

**Copyright:** © 2011 Sperandio et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by PREVENT CD (project: BU-FP6-2005-FOOD4B-contract no. 036883). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: ydonge@unina.it

## Introduction

The pathway of gluten-induced immunoresponse in Celiac Disease (CD) is not yet completely clarified to date, but a tremendous progress was gained recently [1,2] by the discovery of candidate genes identified by genome wide association studies (GWAS) [3–5]. More than 40 candidate genes were discovered, each of them giving a minute contribution to the genomic variance in CD [6]. Only a functional profile of the gluten induced immunoresponse will shed light into the complex interaction of these molecules. The way to progress is to draw reliable pieces of functional pathways that may build up the puzzle of the gluten-induced immunoresponse. New genes, added to HLA, account for around 50% of the genetic risk. Even though relatively homogeneous, CD is probably not one disease, and different pathways and risk factors can be probably distinguished. The progressive stages of the disease are a resource to better understand the role of the genes identified up to date.

Alongside the HLA, a robust association was identified by GWAS on chromosome 4q27, containing the IL-2, IL-21 and KIAA1109

gene cluster [3]. We did confirm this strong association in our CD patients and controls, originated from the same region of Southern Italy (unpublished data). The discovery of a heritable variation in this gene cluster points to a direct involvement of these cytokines. In active CD there is an enhanced production of IL-21 [1]. The quiescence of the active phase of CD, obtained by a gluten free diet (GFD), reverts the IL-21 activity to normal values. When mucosal biopsies from treated CD patients are challenged with gluten peptides, IL-21 is again over expressed [1]. IL-21 induces, in active CD, T-bet expression [7]. The over expression of IL-21 is likely to play a crucial role in the activation of cytotoxic T-cell leading to epithelial cell death and mucosal destruction of active CD [8].

c-REL and TNFAIP3 are key mediators in the nuclear factor kappa B (NF- $\kappa$ B) inflammatory signalling pathway [5]. We suggested previously an early involvement of the NF- $\kappa$ B pathway in the gluten-induced inflammation either in acute CD as well as in gliadin exposed small intestinal mucosa tissue culture [9]. We proposed that, through the NF- $\kappa$ B pathway, gliadin peptides could stimulate iNOS expression in IFN- $\gamma$  treated macrophages from non-celiac patients [10].



So far, genetic studies gave a solid base to explore biological models of gluten-induced immune response, but, no genotype-phenotype association could be found for any of the above listed candidate genes [11].

Recently the clinical presentation of CD underwent substantial changes: from classical malabsorption driven picture to light or absent symptoms associated to moderate damage of intestinal mucosa. The term 'potential CD' was assigned to individuals with proper DQ2 or DQ8, production of anti Transglutaminase (anti TTG) antibodies and normal small intestinal mucosa, classified as Marsh 0 (no damage) or Marsh 1 stage (unspecific intra epithelial infiltration only). These are also defined "serological positive cases" [12–14]. Potential patients suggest that the development of adaptive anti-gluten immunity is not sufficient to develop villous atrophy. This is supported by animal models showing that inflammatory anti-gluten responses are not sufficient to develop mucosal lesions [2]. Thus, we can subgroup individuals into three groups: controls, patients who developed antibodies but no intestinal lesions (M0 and M1 potentials) and patients with antibodies and villous atrophy (overt CD). Analyzing differences among these subgroups may shed light into the genes differentially involved in development of adaptive anti-gluten immunity (different between control and potential) and tissue lesions (different between potential and CD), as well as genes involved in all steps of CD pathogenesis (different between control and potential, and potential and CD).

We have a sizeable cohort of a living experimental model of CD: intestinal production of anti TTG antibodies but no specific mucosal damage. The vast majority of these individuals have no symptoms and may be grouped into low-medium-high anti-TTG producers, but all have a small intestinal mucosa with normal villi/crypt ratio, none or moderate IEL infiltration and normal epithelial layer. Over a prolonged follow up on gluten containing diet only about 1/3<sup>rd</sup> develop mucosal damage, most frequently not associated to symptoms [14]. The reasons why potential CD do not show any degree of substantial mucosal damage, albeit the presence of all the features of CD, is the question, whose answer can contribute to the understanding of the pathogenesis of CD.

Our aim is to explore the presence of genetic and expression factors that may differentiate potential patients from overt CD patients.

## Results

### HLA typing

The HLA genotype of 127 potential CD cases was compared to that of 311 Celiacs previously reported by us [15]. Potential CD less frequently belong to high risk classes (double DQ2 or DQ2 in trans), while they show more frequently very low or moderate HLA-related risk (Table 1). We observed more cases in the HLA class at lower risk among potentials as compared to full celiacs: these cases bear half of the DQ2 heterodimer, either DQB1\*02 or DQA1\*05 only. To avoid a selection bias we did not excluded these potential cases, who produced anti TGASE and were anti Endomysium positive.

When we compared the HLA risk between M0 and M1 potential cases we did not observe any difference ( $\chi^2 = 1.92$ ;  $p = 0.75$ , data not shown). Cases who developed small intestinal atrophy during follow up did not show an HLA distribution different from those who remained potential ( $\chi^2 = 5.17$ ;  $p = 0.27$ , data not shown).

### Eight Candidate Genes SNPs genotype and allele distribution

Hunt et al. [4] reported 8 new CD risk loci, located in different chromosomes. The size of the sample of the potential CD has no

**Table 1.** HLA risk classes distribution.

HLA Haplotype Risk Class	311 CD cases	127 Potential CD cases
Double DQ2	74 (24%)	12 (9.4%)
DQ2 in trans	117 (38%)	23 (18.1%)
DQ2 single	79 (25%)	48 (37.8%)
DQ8 or DQ8*02 (DQA1*05 negative)	13 (4%)	21 (16.5%)
NoDQ2/NoDQ8	28 (9%)	23 (18.0%)
$\chi^2 = 48.42$ $p = 1 \times 10^{-9}$		

doi:10.1371/journal.pone.0021281.t001

sufficient power to differentiate the polymorphisms distribution between potential CD and controls: nevertheless we found that three SNPs in LPP, c-REL and CCR regions respectively showed significant differences between controls and potential CD in their genotype distribution (Table 2). Seven out of 8 SNPs analyzed showed no significant differences between potential and CD cases in their genotype distribution. The SNP marking the c-REL gene showed the GG genotype 2.5 times more often in potentials versus 'full' CD and versus Controls (Table 2).

By analysis of allele distribution, we found that both LPP\*A, c-REL\*G, and CCR\*A alleles were significantly associated to potential cases compared to controls (Table 3). Only c-REL\*G allele was significantly associated with potential cases when compared to CD cases (Table 3).

None of the candidate genes distinguished between M0 and M1 potential cases (data not shown). M0 patients were not different from controls, except for CCR gene. M1 patients reinforced the significant differential distribution of the c-REL genotype when compared to CD: again the GG genotype was about three times more frequent in M1 patients compared to CD patients. The same trend, with good statistical significance, is observed between M1 potential and controls not only for c-REL but also for LPP gene (Table S1).

The eleven cases who developed small intestinal atrophy during follow up did not show an allelic and genotypic distribution different from those who remained potential, maybe because of the sample size that was too small to evaluate differences (Table S2 and S3).

### Five Candidate SNPs on the 4q genomic region

Despite the limited sample size we found that three SNPs in the 4q genomic region showed significant differences between controls and potential CD patients (rs4374642, rs13119723 and rs6840978) in their genotype distribution. In addition, we found also significant differences between potential CD and CD patients for three SNPs (rs4374642, rs13119723 and rs6822844) (Table 4).

The rs4374642, rs13119723 and rs6840978 SNPs, respectively, were significantly associated when Potential CD cases were compared to Controls (Table 5). In contrast, only the rs4374642\*T was the risk allele associated with potential cases when compared to CD cases (Table 5).

Cases who developed small intestinal atrophy during follow up did not show an allelic and genotypic distribution different from those who remained potential (Table S2 and S3).

### Expression Studies

The expression of the IL-21 gene was completely suppressed in potential CD, as compared to CD (Mann-Whitney test; Rank test,

**Table 2.** Association results for 8 celiac non-HLA risk variants.

		Controls	Potential CD	$\chi^2$	p value	CD cases	Potential CD	$\chi^2$	p value
RGS1 (rs2816316)	AA	488 (68.6%)	88 (69.3%)	0.32	0.85	463 (72.7%)	88 (69.3%)	2.67	0.26
	AC	198 (27.1%)	35 (27.6%)			166 (26.1%)	35 (27.6%)		
	CC	30 (4.2%)	4 (3.1%)			8 (1.3%)	4 (3.1%)		
IL18RAP (rs917997)	AA	42 (5.9%)	7 (5.5%)	1.07	0.58	41 (6.4%)	7 (5.5%)	0.69	0.71
	AG	246 (34.6%)	50 (39.4%)			227 (35.6%)	50 (39.4%)		
	GG	423 (59.5%)	70 (55.1%)			369 (57.9%)	70 (55.1%)		
LPP (rs1464510)	AA	108 (15.2%)	28 (22.0%)	7.58	0.02	152 (23.9%)	28 (22.0%)	0.77	0.68
	AC	362 (50.9%)	70 (55.1%)			324 (50.9%)	70 (55.1%)		
	CC	241 (33.9%)	29 (22.8%)			161 (25.3%)	29 (22.8%)		
OLIG3 (rs2327832)	AA	480 (67.7%)	84 (66.7%)	0.05	0.98	399 (63.6%)	84 (66.7%)	0.81	0.67
	AG	202 (28.5%)	37 (29.4%)			208 (32.2%)	37 (29.4%)		
	GG	27 (3.8%)	5 (4.0%)			20 (3.2%)	5 (4.0%)		
TAGAP (rs1738074)	AA	125 (17.6%)	26 (20.5%)	0.6	0.74	144 (22.6%)	26 (20.5%)	0.36	0.84
	AG	354 (49.9%)	61 (48.0%)			305 (47.9%)	61 (48.0%)		
	GG	231 (32.5%)	40 (31.5%)			188 (29.5%)	40 (31.5%)		
c-REL (rs842647)	AA	404 (56.8%)	62 (48.8%)	9.84	0.01	359 (56.4%)	62 (48.8%)	10.8	0.01
	AG	272 (38.3%)	50 (39.4%)			249 (39.1%)	50 (39.4%)		
	GG	35 (4.9%)	15 (11.8%)			29 (4.6%)	15 (11.8%)		
CCR (rs6441961)	AA	99 (13.9%)	25 (19.7%)	8.28	<0.01	120 (18.8%)	25 (19.7%)	2.29	0.32
	AG	318 (44.7%)	66 (52.0%)			293 (46.0%)	66 (52.0%)		
	GG	294 (41.4%)	36 (28.3%)			224 (35.2%)	36 (28.3%)		
SCHIP1 (rs17810546)	AA	610 (85.8%)	102 (80.3%)	2.56	0.28	526 (82.6%)	102 (80.3%)	0.92	0.63
	AG	92 (12.9%)	23 (18.1%)			106 (16.6%)	23 (18.1%)		
	GG	9 (1.3%)	2 (1.6%)			5 (0.8%)	2 (1.6%)		

doi:10.1371/journal.pone.0021281.t002

$p < 0.01$ ) and to controls (Fig. 1A). We underline that the values of expression in potential CD either completely normal as well as infiltrated were consistently lower than the controls ( $p = 0.02$  for both stages). So IL-21 expression is markedly suppressed in these patients. Celiac patients revert completely IL-21 at the levels of controls after at least one year of GFD. But IL-21 expression in celiacs after GFD is still higher than that of the potential CD ( $p = 0.01$ ).

IL-2 is over-expressed in M0 potential CD as compared to controls ( $p = 0.03$ ) and to celiacs on GFD ( $p = 0.01$ ) (Fig. 1B). Fig. 1C shows the expression of the KIAA1109 gene: its pattern is very similar to that of IL-2. It may be noted that M1 cases have an expression 4 times lower than M0 potentials ( $p = 0.04$ ) but a marked increase over the controls ( $p < 0.01$ ) and of about 4 folds over the celiacs ( $p = 0.01$ ). Celiacs on GFD show higher values than acute phase celiacs ( $p = 0.03$ ) and controls ( $p = 0.02$ ).

The expression of the c-REL gene was higher in M0 potentials, compared to CD cases ( $p = 0.04$ ) and to controls ( $p = 0.03$ ) (Fig. 1D).

We explored the relationship between the expression data with the genotype of the same patients for the same gene. We could not recover all DNA samples from the patients, whose RNA was used for expression experiments. Nevertheless we could not find a specific genotype related to the expression value for the same gene, as reported previously [5].

## Discussion

This work attempts to explore the genetic and expression profiles of a special set of individuals in the CD domain. Potential

Celiac patients, with positive serology and undamaged small intestinal mucosa appear to be differently associated with some genetic risk factors compared to full blown CD patients.

Potential CD bears a load of HLA-related CD risk significantly lighter than that of CD cases, with a mix of high-intermediate-low risk different from that observed in CD cases. No difference in HLA profile was observed between M0 and M1 potential patients, neither between the cases who developed full blown disease and those who remained potential at follow up. Subjects without the classical DQ2 or DQ8 heterodimers were double in potential (23, 18.0%) than in classical celiacs (28, 9.0%), but they produced antibodies against Human Tissue Transglutaminase, with positive Endomysial antibodies, just as much as the DQ2 positive cases. In addition we could find no discrimination on the outcome between the DQ2/DQ8 carriers and those who carried only half of the heterodimers. This is one of the results of this work and should not produce a selection bias. Three cases were questionable on the basis of having only DQA1\*01 /DQB1\*0501: but this haplotype was previously described in celiac disease [16]. Although they may not represent the typical potential celiac, we did not exclude them, to avoid selection bias. We better accept that only 3/127 cases may not show the full pattern of the majority of the cohort.

We identified six polymorphisms suggestive to be differently distributed between controls and potential CD: KIAA1109, IL-21, LPP, c-REL and CCR. We suppose that these factors are implicated at some stages of CD pathogenesis, and that, as for HLA, there may be a "gene-dosage effect". Furthermore, c-REL and KIAA1109 are more expressed in M0 potential patients

**Table 3.** Association results of non-HLA SNPs alleles.

Potential CD and controls									
SNP	Locus	A1	A2	Pot CD Cases, Controls Ratios	MAF Pot CD cases	MAF controls	$\chi^2$	p value	Odds ratio (95% confidence interval)
rs2816316	RGS1	A	C	214:44, 1169:253	0.17	0.18	0.08	0.77	1.05 (0.74–1.50)
rs917997	IL18RAP	G	A	198:65, 1092:330	0.25	0.23	0.48	0.49	0.90 (0.66–1.22)
rs1464510	LPP	C	A	129:129, 844:578	0.50	0.41	7.84	0.01	0.68 (0.52–0.89)
rs2327832	OLIG3	A	G	207:40, 1162:256	0.19	0.18	0.17	0.68	0.93 (0.66–1.31)
rs1738074	TAGAP	G	A	143:115, 816:604	0.44	0.42	0.37	0.54	0.92 (0.70–1.20)
rs842647	c-REL	A	G	178:80, 1080:342	0.31	0.24	5.62	0.02	0.70 (0.53–0.94)
rs6441961	CCR	G	A	139:119, 906:516	0.46	0.36	8.99	<0.01	0.66 (0.51–0.87)
rs17810546	SCHIP1	A	G	231:27, 1312:110	0.10	0.08	2.17	0.14	0.72 (0.46–1.12)
Potential CD and CD patients									
SNP	Locus	A1	A2	Pot CD cases, CD cases Ratios	MAF Pot CD cases	MAF CD cases	$\chi^2$	p value	Odds ratio (95% confidence interval)
rs2816316	RGS1	A	C	214:44, 1092:182	0.17	0.14	1.31	0.25	0.81 (0.56–1.16)
rs917997	IL18RAP	G	A	198:65, 965:309	0.25	0.24	0.10	0.75	0.95 (0.70–1.29)
rs1464510	LPP	C	A	129:129, 646:628	0.50	0.40	0.04	0.83	0.97 (0.74–1.27)
rs2327832	OLIG3	A	G	207:40, 1008:248	0.19	0.20	0.05	0.82	1.04 (0.74–1.46)
rs1738074	TAGAP	G	A	143:115, 681:598	0.44	0.46	0.34	0.56	1.08 (0.83–1.42)
rs842647	c-REL	A	G	178:80, 967:307	0.31	0.24	5.43	0.02	0.71 (0.53–0.95)
rs6441961	CCR	G	A	139:119, 741:533	0.46	0.42	1.61	0.20	0.84 (0.64–1.10)
rs17810546	SCHIP1	A	G	231:27, 1158:116	0.10	0.09	0.47	0.49	0.84 (0.55–1.33)

doi:10.1371/journal.pone.0021281.t003

suggesting that they may play a role in the development of adaptive anti-gluten immunity.

The SNP at c-REL gene shows a different distribution of the G allele, as well as of the corresponding genotype, in potential CD versus celiac and versus controls. Unfortunately we were unable to correlate the G allele of the polymorphic marker of the gene with RNA expression, but we will explore further in more cases.

c-REL gene has been recently associated to CD but no significant effect of the rs842647 polymorphism on gene expression was observed [5]. In small intestine tissue from controls and from treated and untreated celiac patients, c-REL gene did not show significant differential expression among the three groups [5]. c-REL, a subunit of the NF- $\kappa$ B complex, is crucial for the regulation of this major nuclear factor for the innate and adaptive immunity

**Table 4.** Association results for 5 celiac non-HLA risk variants, located on 4q27 chromosome.

	Controls	Potential CD	$\chi^2$	p value	CD cases	Potential CD	$\chi^2$	p value
<b>KIAA1109 (rs4374642)</b>	CC		<b>4.52</b>	<b>0.03</b>	4 (1.2%)		<b>7.22</b>	<b>0.03</b>
	CT	63 (14.2%)			53 (15.3%)	9 (7.1%)		
	TT	381 (85.8%)			290 (83.6%)	118 (92.9%)		
<b>KIAA1109 (rs13119723)</b>	AA	397 (88.6%)	<b>5.99</b>	<b>0.05</b>	290 (84.1%)	104 (81.3%)	<b>8.16</b>	<b>0.02</b>
	AG	48 (10.7%)			55 (15.9%)	21 (16.4%)		
	GG	3 (0.7%)				3 (2.3%)		
<b>KIAA1109 (rs1127348)</b>	CC	16 (3.6%)	2.05	0.36	9 (2.6%)	8 (6.3%)	3.51	0.17
	CT	122 (27.5%)			103 (30.3%)	37 (29.1%)		
	TT	205 (68.8%)			228 (67.1%)	82 (64.6%)		
<b>IL2/IL21 (rs6822844)</b>	GG	358 (80.1%)	0.35	0.84	267 (81.9%)	108 (79.8%)	<b>7.61</b>	<b>0.02</b>
	GT	82 (18.3%)			59 (18.1%)	23 (17.8%)		
	TT	7 (1.6%)				3 (2.3%)		
<b>IL21 (rs6840978)</b>	CC	363 (81.6%)	<b>17.94</b>	<b>&lt;0.01</b>	222 (76.8%)	85 (72.0%)	4.5	0.10
	CT	82 (18.4%)			65 (22.5%)	29 (24.6%)		
	TT	4 (3.4%)			2 (0.7%)	4 (3.4%)		

doi:10.1371/journal.pone.0021281.t004



**Table 5.** Association results of non-HLA SNPs Alleles, located on chromosome 4q27.

Potential CD and Controls									
SNP	Locus	A1	A2	Pot CD Cases, Controls Ratios	MAF Pot CD cases	MAF Controls	$\chi^2$	p value	Odds ratio (95% confidence interval)
rs4374642	KIAA1109	T	C	249/9, 823/63	0.03	0.07	4.44	0.03	2.12 (1.04–4.32)
rs13119723	KIAA1109	A	G	229/27, 832/54	0.10	0.06	5.97	0.01	0.55 (0.34–0.89)
rs1127348	KIAA1109	T	C	201/53, 732/154	0.21	0.17	1.61	0.20	0.80 (0.56–1.13)
rs6822844	IL2, IL21	G	T	228/28, 796/96	0.11	0.10	0.01	0.94	0.98 (0.63–1.53)
rs6840978	IL21	C	T	199/37, 802/82	0.16	0.09	8.04	<0.01	0.55 (0.36–0.83)
Potential CD and CD patients									
SNP	Locus	A1	A2	Pot CD Cases, CD cases Ratios	MAF Pot CD cases	MAF CD cases	$\chi^2$	p value	Odds ratio (95% confidence interval)
rs4374642	KIAA1109	T	C	249/9, 622/58	0.03	0.08	7.17	0.01	2.58 (1.26–5.29)
rs13119723	KIAA1109	A	G	229/27, 604/54	0.10	0.08	1.25	0.26	0.76 (0.47–1.23)
rs1127348	KIAA1109	T	C	201/53, 552/120	0.21	0.18	1.10	0.29	0.82 (0.58–1.18)
rs6822844	IL2, IL21	G	T	228/28, 592/58	0.11	0.09	0.87	0.35	0.79 (0.49–1.28)
rs6840978	IL21	C	T	199/37, 508/69	0.16	0.12	1.92	0.17	0.74 (0.48–1.14)

doi:10.1371/journal.pone.0021281.t005

[17]. It's regulation can terminate the NF- $\kappa$ B activation by innate immune stimuli. The integral structure of c-REL is crucial for the activation as well as for the termination of NF- $\kappa$ B induced immunoreponse [18].

In conclusion, the finding of the different allele and genotype distribution of this gene in the potential CD model does point to the crucial role of NF- $\kappa$ B in the pathogenesis of the innate as well as induced immune response in CD. Sustained activation of NF- $\kappa$ B in intestinal mucosa of CD patients leads to prolonged induction of inflammatory genes expression and thereby perpetuates the chronic inflammatory process. Our group produced consistent evidence of the very early involvement of NF- $\kappa$ B mechanism in the gliadin-induced innate immune response in CD [9]. NF- $\kappa$ B is constitutively active in intestinal mucosa of patients with untreated CD and reverts to normal values when gluten is removed from the diet.

The polymorphic markers of the 4q candidate region show significant differences between potential CD and CD cases.

Potential CD show a marked suppression of IL-21, below the level of CD cases, celiacs on GFD and even of normal controls. IL-2 and KIAA1109, on the contrary were both over-expressed in M0 potential CD compared to M1, to celiacs and to controls. M1 cases show an IL-2 and KIAA1109 expression very similar to that of CD.

The finding that the IL-21 gene differentiates potential from controls suggests that it may be implicated also in the adaptive anti-gluten immune responses. This is compatible with findings indicating that IL-21 is produced by follicular helper cells and plays a critical role in B cell differentiation. However IL-21 is also implicated in tissue damage because of its ability to promote NK activity and block regulatory T cells.

In accordance with the role of IL-21 in tissue damage and its down regulation in the mucosa of potential CD patients we found that the IL-21/IL-2 candidate region differentiates potential from CD patients. Conversely, IL-2, which was found to be a critical growth factor for Foxp3+ regulatory T cells, is highly expressed at the transcriptional level in potential CD mucosa. Because Foxp3 regulatory T cells play a critical role in blocking effectors T cell

functions, these finding suggests that the level of IL-2 may be a critical factor in the development of tissue damage.

Potential patients have most of the features of CD patients: they have similar genes, produce the same antibodies but they lack the final destructive phase of the gluten induced immunoreponse, the cytotoxic destruction of small intestinal mucosa. No clinical, serological or histological features could distinguish M0 from M1 patients up to now [14], for the first time we can speculate that potential CD are a mixture of two populations, with different non-HLA genetics and significant differential expression of some candidate genes.

Those potential cases who developed full blown mucosal damage on follow up, did not show any of the time 0 serologic markers worse than those who remained potential over a prolonged follow up. Similarly no genetic polymorphism, including HLA, could predict the outcome of these potential cases. We observed the same distribution of the HLA as well as of the non-HLA genes in the cases who eventually become full blown celiacs.

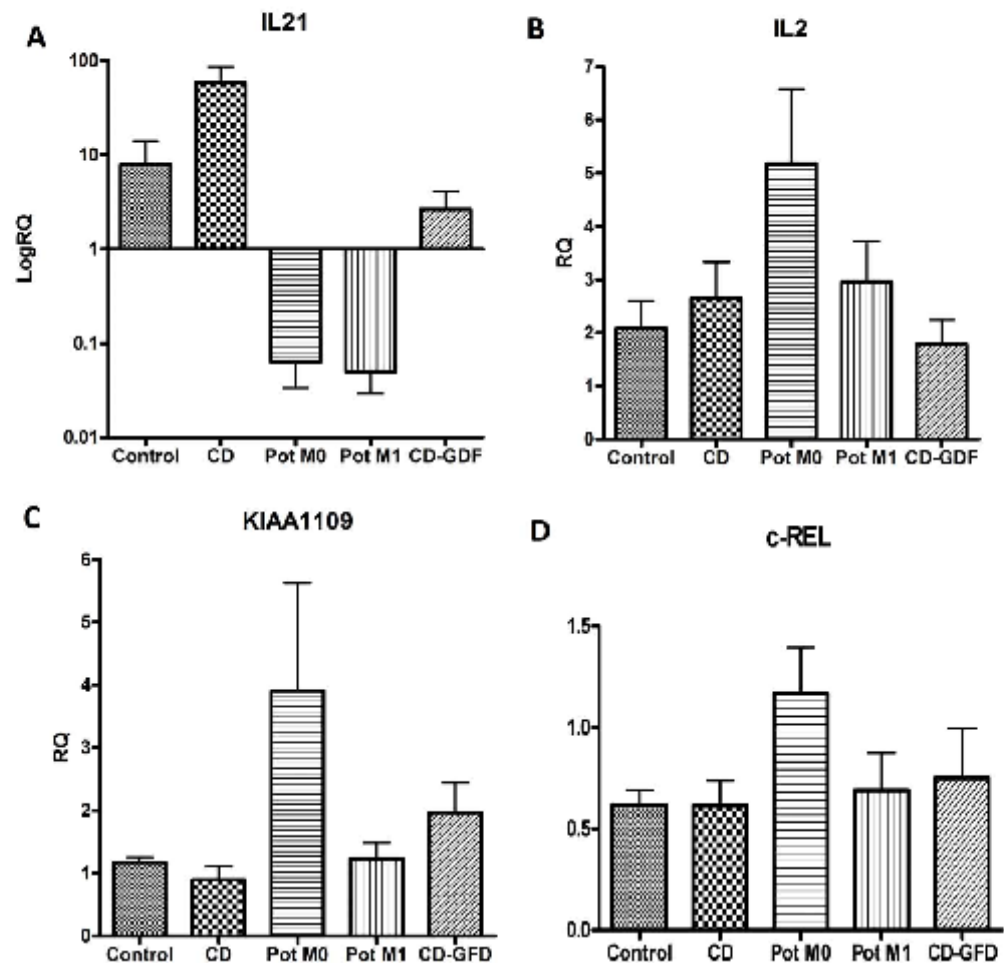
The limit of this study is the relatively short follow up time of this peculiar cohort of Potential CD cases: up to now there has been no possibility to identify any marker (clinical, environmental, serological or genetic) that could help to predict the time of conversion from predisposition to CD to full blown disease. In conclusion we cannot state that the cohort of potential celiacs who have again show an undamaged intestinal mucosa after 4 to 6 years of follow up, will not eventually develop it later in life. We can only observe that this group did not developed the full blown disease after 4–6 years. But the genetic and expression features point to the existence of a peculiar molecular profile, that may help to identify cases who do not have the complete molecular repertoire to develop the disease.

They may indeed be considered a live biological model of CD pathogenesis, where the process is, for some time or even forever, interrupted just before the final destructive phase.

## Materials and Methods

### Ethics statement

The Study Protocol was approved by the Ethics Committee of the University of Naples Federico II. Biopsies were taken during routine



**Figure 1. Expression studies of candidate genes on RNA extracted from small intestinal mucosal biopsies.**  
doi:10.1371/journal.pone.0021281.g001

hospital admission requested for diagnostic purposes: parents or guardian gave their informed consent about the endoscopic procedures and the biopsy. Patients did not underwent blood or specimen sampling over and above those requested by the routine diagnostic procedures according to the ESPGHAN guidelines [19].

#### Subjects

1. 127 children (76 females, median age 6 years and 6 months, range 18 months–16 years) were classified as 'potential CD' on the basis of:

- Increased levels of anti-TTG (IgA anti Human Tissues Transglutaminase) and Anti-Endomysium positive. Serum

EMA and anti TTG IgA were detected by indirect immunofluorescence and by enzyme-linked immunosorbent assay (ELISA) technique using a kit based on human recombinant antigen respectively (Eu-tTg IgA Kit, Eurospital, Trieste).

- Architecturally normal small intestinal mucosa (M0 and M1) on at least four mucosa samples, to minimize the possible bias due to patchy lesions. Definition of the stage of lesion was obtained by Marsh stages modified by Oberhuber [14].

At the first biopsy, patients underwent clinical and laboratory assessment (CD-related auto antibodies, thyroid auto antibodies, nutritional parameters, bowel inflammatory indexes). 50/127 (39.4%) patients belong to at risk groups: 32 were first-degree

relatives of celiac patients, 18 were affected by autoimmune diseases (11 type 1 diabetes and 7 thyroiditis). 105/127 (82.7%) were asymptomatic; 21/127 showed symptoms attributable to CD, 17 gastrointestinal (abdominal pain, weight loss, diarrhoea, failure to thrive) and 4 extra intestinal (2 short stature, one dilatative cardiomyopathy, one refractory anaemia). 21 symptomatic patients went on GFD, the others continued on a normal diet with a 6 monthly follow-up schedule. After two years they were re-biopsied. The second biopsy was anticipated if symptoms ensued.

The 21 patients on a GFD became negative for CD-related autoantibodies but 6 of them did not show any clinical response to GFD. 105/127 patients continued a gluten containing diet. During follow-up CD serology became negative in about 15% and antibody fluctuation was observed in 42% with transient negativity for anti-TTG values. Four patients developed autoimmune disorders (1 type 1 diabetes and 3 thyroiditis), 1 patient developed a vitamin K deficiency and began a GFD. 44/105 patients repeated a biopsy after 2 years of follow-up. 13/44 (29.5%) developed mucosa damage, all the others were invariant at the control biopsy [14]. In conclusion, 35 children were considered celiac patients: 23 patients were put on a GFD at the beginning of the study because of symptoms, 1 patient was put on a GFD due to appearance of symptoms during the follow-up and 11 patients developed atrophy.

- To evaluate genetic risk allele a cohort of 643 overt CD cases, and 711 healthy controls were compared to the 127 potential CD patients. Controls were randomly selected from a DNA bank representing the healthy population of the region, stratified for province. All originated from the same centre and the same geographical area. Genotyping of the 4q candidate gene was performed on a set of about 350 cases and about 450 controls from the above sample.
- To evaluate the HLA-related risk the 127 potential cases were compared to a set of 311 overt cases from the same sample described above and reported in Bourgey et al. [15].

## Genotyping

HLA typing was performed using Eu-Gene Risk kit (EURO-SPTAL). Cases were grouped into 5 HLA, as reported previously [15]. Thirteen SNPs were genotyped using TaqMan technology

## References

- Morise B, Ripoch J, Heyman M, Gelfand-Benjamin N (2009) Celiac disease: from oral tolerance to intestinal inflammation, autoimmunity and lymphoma-genes. *Nature Rev* 2: 8–23.
- Jabri B, Solid LM (2009) Tissue-mediated control of immunopathology in celiac disease. *Nat Rev Immunol* 9: 858–870.
- van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, et al. (2007) A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 39: 827–829.
- Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, et al. (2008) Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 40: 395–402.
- Trynka G, Zhernakova A, Romanos J, Franke L, Hunt KA, et al. (2009) Celiac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. *Gut* 58: 1078–1083.
- Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, et al. (2010) Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 42: 295–302.
- Fina D, Sarin M, Caruso R, Del Vecchio Blanco G, et al. (2008) Interleukin 21 contributes to the mucosal T helper cell type 1 response in celiac disease. *Gut* 57: 887–892.
- Strengell M, Jalkanen I, Mäkeläinen S (2003) IFN- $\alpha$  regulates IL-21 and IL-21R expression in human NK and T cells. *J Leukoc Biol* 75: 416–422.
- Maiuri MC, De Stefano D, Mele G, Fecarotta S, Greco I, et al. (2008) Nuclear factor kappa B is activated in small intestinal mucosa of celiac patients. *J Mol Med* 81: 373–379.
- Maiuri MC, De Stefano D, Mele G, Iovine B, Bontade MA, et al. (2003) Gliadin increases iNOS gene expression in interferon-gamma-stimulated RAW 264.7 cells through a mechanism involving NF-kappa B. *Nutrition* 19: 63–71.
- Greco I, Percepo S, Clot F, Bouguerra F, Baham MC, et al. (1998) Lack of correlation between genotype and phenotype in celiac disease. *J Pediatr Gastroenterol Nutr* 26: 286–290.
- Kokkinou O, Collin P, Kosman-Schoen I, Sahni T, Iliadis S, et al. (2008) Gluten-dependent small bowel mucosal transglutaminase 2-specific IgA deposits in overt and mild enteropathy celiac disease. *J Pediatr Gastroenterol Nutr* 47: 436–442.
- Kuppa K, Ashorn M, Iliadis S, Kokkinou O, Savolainen P, et al. (2010) Celiac Disease without Villous Atrophy in Children: A Prospective Study. *J Pediatr* 157: 373–380, 380.e1.
- Toes A, Salvati VM, Auricchio R, Maglio M, Borrelli M, et al. (2010) Most children with potential celiac disease are healthy but one-third of them develop villous atrophy at 3 years follow-up. *Clin Gastroenterol Hepatol* 17. [Epub ahead of print].
- Bourgey M, Calcagno G, Tinto N, Genovese D, Marguerite-Jeanin P, et al. (2007) HLA related genetic risk for celiac disease. *Gut* 56: 1054–1059.

(Applied Biosystems) as described elsewhere [3,20]. Reactions were performed on 7900HT Fast Real-Time PCR System (Applied Biosystems).

## Expression Studies

The expression of KIAA1109, IL-2 and IL-21 was analyzed in mucosal biopsy samples from 17 CD cases, 21 potential CD patients (7 M0 and 14 M1), 8 CD patients on GFD, and 14 controls using Real Time PCR by 'Assays-On-Demand' (Applied Biosystems). Controls underwent endoscopy for functional disorders, from these we excluded subjects with *Helicobacter pylori* infection, known to modify the expression of IL-21 [21]. The analysis of the expression of c-REL gene was performed in mucosal biopsy samples from 10 CD cases, 10 potential CD patients (5 M0 and 5 M1), and 10 controls.

## Statistical Analysis

Frequencies were compared by the  $\chi^2$  test with 0.05 probability of the null hypothesis. Skewed variables were log transformed when required. Expression data were compared by a signed Rank test for independent samples. Data analysis was performed using: Prism (GraphPad Software, San Diego, CA), SPSS 15 (SPSS Inc., Chicago, IL, USA) and Haploview 4.1.

## Supporting Information

**Table S1 Comparison between cases with completely normal mucosa (M0) and those with infiltrated mucosa (M1).**  
(DOCX)

**Table S2 Association results of 13 non-HLA SNPs alleles.**  
(DOCX)

**Table S3 Association results for 13 celiac non-HLA risk variants.**  
(DOCX)

## Author Contributions

Conceived and designed the experiments: RT SA LG. Performed the experiments: MPS VI FT. Analyzed the data: MPS VI FT LG. Contributed reagents/materials/analysis tools: AT RA JR GT. Wrote the paper: MPS AT VI FT LG. Critical manuscript revision: RT BJ RT SA.

16. Karel K, Leuka AS, Mondie SJ, Acher H, Clot F, et al. (2003) European Genetics Cluster on Celiac Disease. *Hum Immunol* 64: 469–477.
17. Bonizzi G, Karin M (2004) The two NF- $\kappa$ B activation pathways and their role in innate and adaptive immunity. *Trends Immunol* 25: 280–288.
18. Lawrence T, Behm M, Liu GV, Nizet V, Karin M (2005) IKK $\alpha$  limits macrophage NF- $\kappa$ B activation and contributes to the resolution of inflammation. *Nature* 434: 1138–1143.
19. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition (1990) Revised criteria for the diagnosis of celiac disease. *Arch Dis Child* 65: 909–911.
20. Romanos J, van Diemen CC, Nolte IM, Trynka G, Zhemakova A, et al. (2009) Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. *Gastroenterology* 137: 834–840.
21. Camas R, Fina D, Pokuo I, Fantini MC, Tosti C, et al. (2007) IL-21 is highly produced in *Helicobacter pylori*-infected gastric mucosa and promotes gelatinase synthesis. *J Immunol* 178: 5957–5965.



# Improving the Estimation of Celiac Disease Sibling Risk by Non-HLA Genes

Valentina Izzo<sup>1,2</sup>, Michele Pinelli<sup>3</sup>, Nadia Tinto<sup>4,5</sup>, Maria Valeria Esposito<sup>4,5</sup>, Arturo Cola<sup>4,5</sup>, Maria Pia Sperandio<sup>1,2</sup>, Francesca Tucci<sup>1,2</sup>, Sergio Cocozza<sup>3</sup>, Luigi Greco<sup>1,2\*</sup>, Lucia Sacchetti<sup>4,5</sup>

**1** Department of Pediatrics, University of Naples Federico II, Naples, Italy, **2** European Laboratory for Food Induced Diseases, University of Naples "Federico II", Naples, Italy, **3** Department of Cellular and Molecular Biology and Pathology "L. Califano", University of Naples "Federico II", Naples, Italy, **4** CEINGE Advanced Biotechnology, S.p.A., Naples, Italy, **5** Department of Biochemistry and Medical Biotechnology, University of Naples "Federico II", Naples, Italy

## Abstract

Celiac Disease (CD) is a polygenic trait, and HLA genes explain less than half of the genetic variation. Through large GWAS more than 40 associated non-HLA genes were identified, but they give a small contribution to the heritability of the disease. The aim of this study is to improve the estimate of the CD risk in siblings, by adding to HLA a small set of non-HLA genes. One-hundred fifty-seven Italian families with a confirmed CD case and at least one other sib and both parents were recruited. Among 249 sibs, 29 developed CD in a 6-year follow-up period. All individuals were typed for HLA and 10 SNPs in non-HLA genes: CCR1/CCR3 (rs6441961), IL12A/SHIP1 and IL12A (rs17810546 and rs9811792), TAGAP (rs1738074), RGS1 (rs2816316), LPP (rs1464510), OLIG3 (rs2327832), REL (rs842647), IL2/IL21 (rs6822844), SH2B3 (rs3184504). Three associated SNPs (in LPP, REL, and RGS1 genes) were identified through the Transmission Disequilibrium Test and a Bayesian approach was used to assign a score (BS) to each detected HLA+SNPs genotype combination. We then classified CD sibs as at low or at high risk if their BS was respectively < or ≥ median BS value within each HLA risk group. A larger number (72%) of CD sibs showed a BS ≥ the median value and had a more than two fold higher OR than CD sibs with a BS value < the median (O.R.= 2.53, p=0.047). Our HLA+SNPs genotype classification, showed both a higher predictive negative value (95% vs 91%) and diagnostic sensitivity (79% vs 45%) than the HLA only. In conclusion, the estimate of the CD risk by HLA+SNPs approach, even if not applicable to prevention, could be a precious tool to improve the prediction of the disease in a cohort of first degree relatives, particularly in the low HLA risk groups.

**Citation:** Izzo V, Pinelli M, Tinto N, Esposito MV, Cola A, et al. (2011) Improving the Estimation of Celiac Disease Sibling Risk by Non-HLA Genes. PLoS ONE 6(11): e26920. doi:10.1371/journal.pone.0026920

**Editor:** Yun Li, University of North Carolina, United States of America

**Received:** July 4, 2011; **Accepted:** October 6, 2011; **Published:** November 7, 2011

**Copyright:** © 2011 Izzo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was funded by the PREVENT-CD project: EU-FP6-2005-FOOD4B-contract no. 016383 and by CBNGE Regione Campania (DGR 1901/2009) Italy. The authors thank Associazione Italiana Celiachia (AIC) for a supporting grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: ydongre@unina.it

## Introduction

Celiac disease (CD) is a chronic small-intestinal enteropathy, triggered by gluten proteins contained in wheat, barley and rye [1]. The evidence of a strong genetic component is suggested by a remarkable familial aggregation: the prevalence of CD is, in fact, 10 times higher in first degree relatives (~10%) than in the whole population (1%) [1–3] and a very high concordance (>80%) is found in monozygotic twins [4]. CD prevalence increased significantly in the last 20 years, so becoming a major public health problem, for this reason, in the near future the CD families are predicted to be a major source of new cases and consequently the CD risk prediction in these cohorts may be important.

At present Histocompatibility Leucocyte Antigens (HLA) explains ~35% [5] of the genetic variance associated to CD. We previously graded the HLA risk genotype in 5 risk groups (from G1 to G5) and we were able to calculate the risk in each group with very wide confidence intervals (from 0.1% to ≥20%) [1]. In particular the higher risk groups (>10%) were those belonging to G1 and G2 groups [1]. However, since HLA alone can explain about 1/3rd of the genetic susceptibility to the disease, other variants should be implicated.

In the last four years, several Genome Wide Association studies (GWAS) identified about 40 genomic regions harboring 64 candidate genes, which are involved in adaptive and innate immunity in CD and also linked to other autoimmune diseases [6–13]. Unfortunately altogether non-HLA genes account for only 4% of the genetic variance [13].

In the field of complex diseases, particularly in CD, great attention is now paid to use the available genetic data to predict the risk of disease in asymptomatic individuals and to support the diagnosis in difficult cases. Although it was described that non-HLA genes improve the ability to identify individuals at high risk, the increased predict ability by only genetics seems still modest in the general population [14]. However the use of non-HLA genes in the disease risk prediction in CD sibs has not yet been explored.

The Bayesian approach was shown to be useful in the managing the GWAS results, for example, in predicting the susceptibility to breast cancer [15]. Applying the same approach in a CD family cohort, we wish to improve the estimation of CD risk among siblings of Celiacs over the available risk HLA-based and thus provide a better tool to evaluate the health status or to predict the disease in these at risk individuals.



## Results

### SNPs evaluation and computation of the Bayesian Score

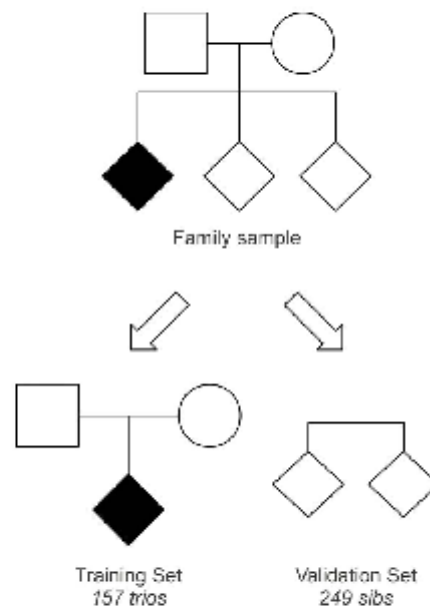
All individuals were typed for 10 CD previously associated SNPs [13] and the TDT test was performed on results obtained from 157 trios of the training set (Fig. 1). Three out of ten investigated SNPs (rs1464510 in LPP, rs842647 in REL and rs2816316 in RGS1 genes) were significantly associated with CD (Table 1). In particular, rs1464510 in LPP gene showed a strong association ( $p < 0.001$ ) according to an additive model, whereas, both rs2816316 in RGS1 and rs842647 in REL genes were also significantly associated (respectively  $p = 0.025$  and  $p = 0.034$ ), by a recessive model. For supporting information about allelic and genotypic frequencies observed in the sample see Table S1 and S2.

In order to evaluate the occurrence of HLA-SNP interaction, we stratified the training set (Fig. 1) according to the HLA risk group of the proband. No statistically significant interaction was found between HLA and non-HLA genes (data not shown).

To compute the Bayesian Score (BS) we compared the frequency of each HLA+SNPs genotype combination detected in probands and in controls (Training set, Fig. 1). Through this approach we obtained a BS for each HLA+SNPs genotype combination (data not shown).

### Validation of the BS and testing of a classification model

The validation set (Fig. 1) was composed by the sibs of the probands, both affected ( $n = 29$ ) and unaffected ( $n = 220$ ). In these subjects we evaluated if HLA+SNPs genotyping could improve the identification of CD risk in sibs better than HLA only.



**Figure 1. Study design of CD families.** The family set was splitted in a Training set, that is 157 trios composed by the 157 probands and their unaffected parents and in a Validation set, that is 249 sibs of the probands.  
doi:10.1371/journal.pone.0026920.g001

**Table 1. Genotypic Transmission Disequilibrium Test (TDT) results.**

SNP	Gene	Model	Risk Allele	OR (95% CI)	p value
rs1464510	LPP	Additive	A	2.36 (1.64–3.41)	<0.001
rs2816316	RGS1	Recessive	A	1.75 (1.07–2.86)	0.025
rs842647	REL	Recessive	A	1.66 (1.04–2.65)	0.034
rs2327832	OU3	Additive	G	1.35 (0.90–2.03)	0.150
rs6441961	CCR1/CCR3	Additive	A	1.24 (0.89–1.72)	0.189
rs6822844	IL2/IL21	Additive	C	1.43 (0.82–2.49)	0.210
rs1738074	TAGAP	Dominant	A	1.31 (0.79–2.16)	0.298
rs3184504	SH2B3	Additive	A	1.19 (0.86–1.63)	0.294
rs17810546	IL12A	Additive	G	1.10 (0.80–1.51)	0.572
rs9811792	IL12A/SCHIP1	Dominant	G	1.10 (0.59–2.05)	0.753

For supporting information about allelic frequencies observed in Trios (Probands and unaffected parents) and in sibs (affected and unaffected) see Table S1.

doi:10.1371/journal.pone.0026920.t001

We assigned to each sib, both affected and unaffected, the BS value corresponding to his HLA+SNPs genotype combination as previously calculated in the training set (Table 2). We observed an increase of average BS values from HLA group 5 to HLA group 1 and, within all 5 HLA groups we found an increase of the BS corresponding to an increase of "A" alleles in the haplotype combination. In order to identify sibs at high risk to develop CD, we distributed affected and unaffected sibs on the basis of their BS (under or above median BS) within each HLA group from G1 to G5 (Fig. 2). Considering the distribution of all the sibs, it is evident that above the median BS within each HLA group, there was always a larger number of affected sibs than under the median: 72% (21/29) versus 28% (8/29). Interestingly, 2/29 affected sibs that being at lower HLA risk (HLA group 4 and 5) could be misclassified, were correctly classified by their BS above the median. We calculated the Odds Ratio (OR) of CD classification based on the BS, above or under the median, within each HLA group. CD sibs with a BS value above the median, had more than two fold higher risk (OR) compared to CD sibs with a BS value under the median (2.53, 95% C.I.: 1.68–3.39;  $p = 0.047$ ).

In our previous work, we estimated the HLA related risk to develop CD, defining a risk range from 0.01 to 0.21 [1]. To refine the previous HLA based estimation of risk by this new approach we set the HLA related mean risk as the *a priori* risk of the Bayesian Model. Having set the HLA risk at the median level of the Bayesian Score, we produced a picture of the variation of risk given by the non-HLA genes for each HLA class (Fig. 3). It is remarkable that the addition of the 3 SNPs does modify the HLA only risk through the 5 HLA risk classes, also in subjects with no DQ2 neither DQ8 haplotype, where the increase in risk is significantly larger than that observed in HLA risk classes.

The addition of only 3 SNPs to the HLA significantly improved the prediction of CD risk in sibs, identifying, within a specific HLA group, those individuals which are more likely to become celiacs. In fact, considering HLA 1 and 2 as the highest risk groups [1], we calculated the diagnostic characteristics of our proposed HLA+SNPs genotype combination obtaining both an higher predictive negative value (NPV) and an higher diagnostic sensitivity (DS) than HLA only, respectively 95% vs 91% (NPV) and 79% vs 45% (DS) (Table 3). Although the discovery of 39 polymorphism associated to CD improved the estimation of the

**Table 2.** Bayesian Score (BS) assigned to each HLA-SNPs genotype combination.

Associated SNPs			BS				
LPP (rs1464510)	REL (rs842647)	RGS1 (rs2816316)	HLA Group 1	HLA Group 2	HLA Group 3	HLA Group 4	HLA Group 5
CC	AG GG	AC CC	0.66	0.63	0.32	0.30	0.04
CC	AG GG	AA	0.74	0.73	0.42	0.39	0.06
CC	AA	AC CC	0.71	0.69	0.38	0.35	0.05
CC	AA	AA	0.79	0.77	0.48	0.46	0.08
AC	AG GG	AC CC	0.81	0.80	0.52	0.49	0.09
AC	AG GG	AA	0.87	0.86	0.62	0.60	0.13
AC	AA	AC CC	0.85	0.84	0.58	0.55	0.11
AC	AA	AA	0.90	0.89	0.68	0.66	0.16
AA	AG GG	AC CC	0.91	0.90	*	0.69	0.18
AA	AG GG	AA	0.94	0.93	0.79	0.77	0.25
AA	AA	AC CC	0.93	0.92	0.76	0.74	0.22
AA	AA	AA	0.95	0.95	0.83	0.81	0.30

\*The combination LPP\*AA - REL\*AG|GG - RGS1\*AC|CC for HLA group 3 was not found among our sibs' cohort.  
doi:10.1371/journal.pone.0026920.t002

heritability of only 4–5% [13], in this Bayesian model the addition of only three SNPs of associated genes improves the sensitivity of risk prediction of 34% compared to the HLA only model.

The computation of the AUC of the ROC curve output a C statistic equal to 0.70 for HLA and 0.73 for HLA+SNPs classifications, showing that the inclusion of SNPs moderately improved the prediction ability.

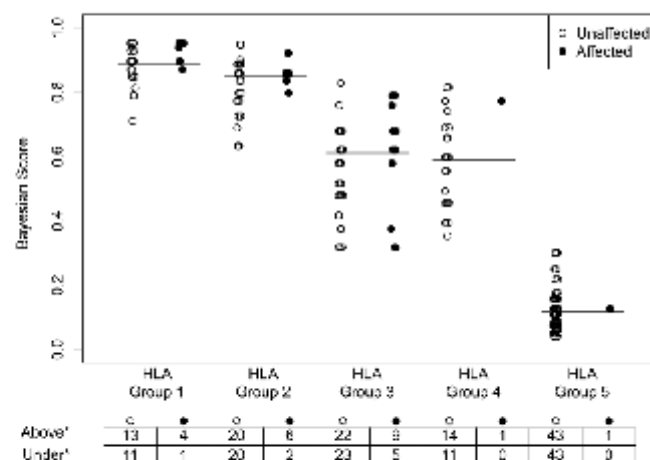
## Discussion

In our previous work, we considered in CD families the risk to develop the disease according to a specific HLA haplotypes,

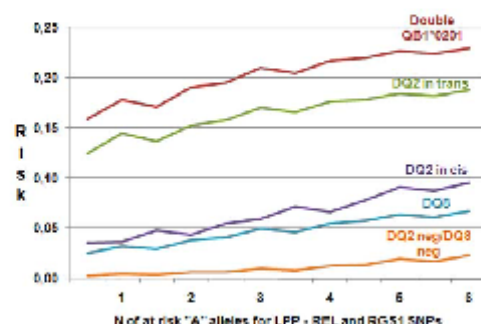
obtaining a risk range from 0.01 to  $\geq 0.20$  [1]. In the present study we evaluated the role of 3 non-HLA genetic markers to influence the CD risk in first relatives of CD affected children.

We collected data on families with at least one CD-affected among offspring. This family set helped to evaluate the association between SNPs and CD (TDT on parents-offspring trios) and to estimate the risk of CD in the other sibs. The TDT design provides robustness to population stratification and mitigation of the possible confounding effect of environmental factors, because all family members share the same environment [16].

Ten SNPs, selected from those previously found to be associated with CD by GWAS [13], were successfully genotyped. In our



**Figure 2.** Distribution of CD sibs based on their BS. Affected ( $n=29$ ) and unaffected ( $n=220$ ) CD sibs were classified on the basis of their BS value  $<$  or  $\geq$  median BS within each HLA group. Horizontal lines correspond to the median BS of all sibs in each HLA risk-group (HLA Group 1=0.90, HLA Group 2=0.86, HLA Group 3=0.62, HLA Group 4=0.60, HLA Group 5=0.13).  
doi:10.1371/journal.pone.0026920.g002



**Figure 3. Refining the CD risk estimate.** The picture shows the modification of the *a priori* HLA related risk by the number of the at risk 'A' alleles of the LPP, REL and RGS1 SNPs. From top to bottom lines correspond to HLA group 1 to 5.  
doi:10.1371/journal.pone.0026920.g003

population three SNPs resulted significantly associated with CD (those in LPP, RGS1 and REL genes) and the other seven investigated SNPs, even if not statistically associated with CD, showed always a higher frequency of the previously reported risk alleles [13] in affected subjects than in controls.

The three genes selected appear to be appealing for the pathogenesis of CD. LPP (OR = 2.36;  $p < 0.001$ ) was reported to be highly expressed in small intestinal mucosa and may have a structural role at sites of cell adhesion in maintaining cell shape and motility [7]. RGS1 (OR = 1.75;  $p = 0.025$ ) belongs to a family of RGS genes. It attenuates the signaling activity of G-proteins, blocking the homing of Intra Epithelial Lymphocytes (IELs), and it is specifically expressed both in human small intestinal mucosa and in murine IELs, key players in the development of human CD villous atrophy [7,17]. REL (OR = 1.66;  $p = 0.034$ ) is a subunit of NF- $\kappa$ B complex, implicated in T cell differentiation [18] and it appears to be a key molecule regulating inflammation and the switch from tolerance to autoimmunity [19]. It is interesting to note that our data confirm previous pathogenetic implications reported in literature of these SNPs with CD as well as with other autoimmune diseases [20].

By the Bayesian approach we calculated a ranking score (BS) among the sibs. However, it should be considered that BS is not a plain disease risk, rather a method to rank different genotypes according to their contribution to make an individual susceptible to CD. For instance, some of our BS are very near to 1, nevertheless none of the considered genotypes could give a 100% risk to develop the disease. In other terms, we considered the BS as a ranking measure, only stating that a given genotype could assign a higher risk than another genotype but does not allow a quantitative measure of the risk difference (2-fold, 3-fold, etc.). However, even if the addition of only 3 SNPs to HLA could be considered at "minor effect" [13], we demonstrated that they could significantly improve the prediction of CD risk in sibs, in terms of diagnostic sensitivity and negative predictive value. So, in a cohort of CD families, our data confirm that non-HLA SNPs evaluation is an useful diagnostic tool in CD risk evaluation as a previous study showed in CD unrelated subjects [14].

CD, on the basis of the actual knowledge, cannot be exactly predicted by genetic testing, but a reliable probabilistic method might be associated to careful surveillance of infants carrying the higher risk. This will help to significantly reduce the heavy load of

**Table 3. Diagnostic characteristics of HLA and HLA-SNPs genotype combination.**

	HLA risk groups 1–2	HLA-SNPs genotype combination
Sensitivity	0.45 (0.27–0.64)	0.79 (0.73–0.84)
Specificity	0.71 (0.65–0.77)	0.54 (0.48–0.60)
NPV	0.91 (0.87–0.96)	0.95 (0.92–0.98)
PPV	0.17 (0.12–0.22)	0.19 (0.14–0.24)

doi:10.1371/journal.pone.0026920.t003

anxiety and pain associated with the appearance of symptoms of CD, by anticipating, with simple serological tests, the clinical appearance of the disease.

To improve the possibility to identify high risk patients in CD families we propose in alternative to the classical HLA classification (Fig. 4, panel A) a slight improved flow-chart (Fig. 4, panel B): 1) HLA genotyping: subjects belonging to the HLA risk groups 1 and 2 will be classified as at high CD risk; 2) subjects belonging to the HLA risk groups 3 and 4, will be further investigated for our SNPs combination (LPP, REL, RGS1) in order to calculate their BS (Fig. 4, panel B). Among these latter subjects those with a BS  $\geq$  the median value will be classified at high risk; 3) subjects belonging to the HLA risk group 5 will be considered at low CD risk. All CD families belonging to the above high risk groups (HLA group 1–2 and HLA group 3–4 with BS  $\geq$  median) will be undergo a strict surveillance.

One of the limitation of our cohort family study could be the sample size, which may have not allowed to explore genes at smaller effect, so explaining the lack of association between SNPs in TAGAP, IL2/IL21, OLIG3, OCR, SH2B3, IL12A and IL12A/SCHIP1 genes with CD although the trend observed in previous studies in unrelated CD patients was confirmed [13]. In the main time the homogeneity of the genetic and environmental domains in the tested families allows to explore risk factors within a controlled cohort. A second limit of the study is the relatively short (6 years) follow up of the sibship, which could cause an underestimation of the disease development at later ages. Our aim is to go on with the monitoring of these families in the next years.

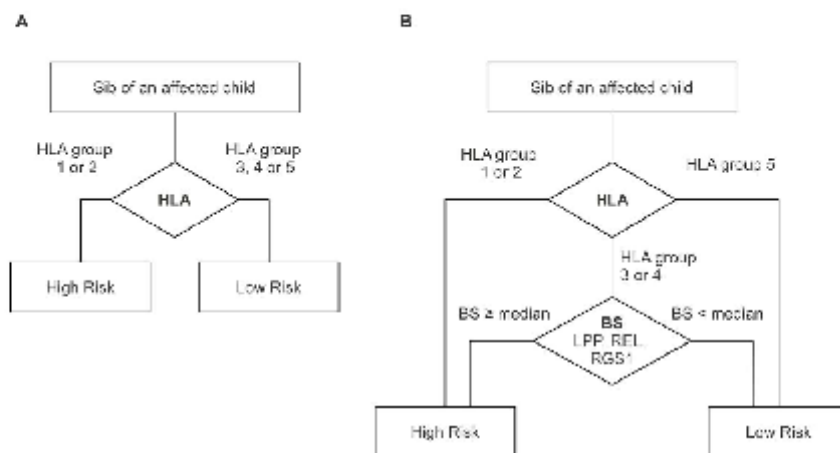
In conclusion, the estimate of the CD risk by HLA+SNPs approach, even if not applicable to prevention, could be a precious tool improving CD diagnosis respect to the only HLA (NPV: 95% vs 91%, and DS: 79% vs 45%), in the cohort of first degree relatives. In fact in clinical practice the absence of HLA risk groups 1 or 2, allows to exclude the disease with high probability, while testing the three SNPs in HLA groups 3 or 4 could represent a further tool to identify less frequent CD cases. So, an infant with high HLA+SNPs score even if belonging to HLA low risk groups, shall undergo a simple surveillance system to allow proper diagnosis and treatment before the full blow disease appears.

## Materials and Methods

### Ethical Statement

The written informed consent was obtained from all participants and from both parents for children. The research was approved by the Ethics Committee of the School of Medicine, University of Naples "Federico II" and was according to principles of the Helsinki II declaration.





**Figure 4. Classification flow-chart.** In panel A the classical HLA-based classification. In panel B the proposed BS-based classification considering the genotypes of HLA plus LPP, RGS1 and REL SNPs.  
doi:10.1371/journal.pone.0026920.g004

#### Family Samples

A cohort of CD families was recruited as previously described [1]. Families included a symptomatic CD patient (hereafter referred as the proband), both parents and at least one sib (for a total of 183 probands, 366 parents and 249 sibs); all probands, as well as the new cases, were diagnosed according to the European Society of Paediatric Gastroenterology and Nutrition (ESPGHAN) criteria [21]. Among the 249 sibs, 29 resulted to be affected over a 6 years follow up program [1].

All individuals were grouped into five decreasing risk classes according to their HLA genotype: very high (>20% with two copies of DQ2.5, or DQ2.5/DQ2.2 Group 1), high (15–20% with DQ2.2/DQA105, Group 2), intermediate (10–15% with one copy of the DQ2.5 heterodimer, Group 3), moderate (1–10% with a double copy of DQ8 or DQ2.2/DQ8, or double copy of DQ2.2, Group 4) or negligible (<1%, with other genotype, Group 5) (Table S3) [1].

#### Non-HLA Single Nucleotide Polymorphisms (SNP) typing

The 798 patients were genotyped for 10 non-HLA SNPs associated with CD: rs6441961 on 3p21 (*Chemokine C-C motif receptor 1 and 3 – CCR1/CCR3*), rs17810546 and rs9811792 on 3q25-26 (*SCHIP1 – Schizophrenia interacting protein 1 – and IL12A – Interleukin 12A*), rs1738074 on 6q25 (*T cell activation GTPase activating protein – TAGAP*), rs2816316 on 1q31 (*Regulator of G-protein signaling 1 – RGS1*), rs1464510 on 3q28 (*Lipoma Preferred Partner – LPP*), rs2327832 on 6q23.3 (*Oligodendrocyte transcription factor 3 – OLIG3*), rs842647 on 2p16.1 (*Reticuloendotheliosis virus oncogene homolog – REL*), rs6822844 on 4q27 (*Interleukin 2 and 21 – IL2/IL21*), rs3184504 on 12q24 (*SH2B adaptor protein 3 – SH2B3*). Genotyping reactions were performed using TaqMan® SNP Genotyping Assays on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA); the final volume was 5 µL, containing master mix, TaqMan assays and about 60 ng of genomic DNA template. All 384 well plates were filled using Biomek® FX (Beckman Coulter, Indianapolis, IN, USA). Allelic Discrimination results were analyzed through the SDS software ver. 2.3.

#### Analysis strategy and statistics

In order to develop the model we split the sample into a training and a validation set (Fig. 1). In the *training set*, composed by probands and their unaffected parents, called *trios*, we evaluated which SNP was associated to the disease, independently from the HLA haplotype. Twenty-six CD families were excluded because they did not meet the requirements for the analysis (e.g. there was an affected parent or patients had an incomplete genotyping), so the training set was composed by 157 trios (Fig. 1). As control haplotypes we considered the haplotypes carried by parents and not transmitted to the affected probands, as they could be representative of the haplotypes in the general population, from which cases are originated. The frequencies of HLA and non HLA haplotypes in controls were estimated by the AFBAC (Affected Family Based Controls) method implemented in the MASC software tool [22]. To evaluate the association of the SNPs with the disease, we performed a *Transmission Disequilibrium Test* (TDT) [16] by using the *trio* package ver. 1.1.15 for R statistical computing software ver. 2.11.2. This method is based on multivariate logistic regression. We considered: dominant, the heterozygote genotype when it had the same risk effect of the high risk homozygote genotype; recessive, the heterozygote genotype when it had the same risk effect of the low risk homozygote genotype; additive, the heterozygote genotype when it had a risk effect intermediate between the two homozygote genotypes. Firstly, we evaluated monovariate association between each SNP and the disease. Secondly, among those resulted to be significantly associated with CD, we evaluated for SNP-SNP interactions by logistic regression model. Finally, the interaction between significant SNPs and HLA groups was verified. In details, we repeated association statistics between CD risk and SNPs on different strata of HLA groups and, thus, computed the interaction tests.

To evaluate the joint effect of HLA and SNPs to influence the risk to develop CD in sibs, we implied a Bayesian approach, focusing only on those SNPs resulted significantly associated with CD at TDT analysis. Usually Bayes' revision of probabilities

allows the computation of an individual probability of an event given the *a priori* data from the general population where the patient comes from. We arbitrarily selected an *a priori* probability of 0.5, because in our situation we could not apply the Bayes' revision of probabilities and obtain an *a posteriori* risk of CD in sibs. Indeed the estimated 10% of CD risk in first grade relatives also account for the role of the genetics, therefore it cannot be used as an *a priori* probability to calculate the *a posteriori* probability, after considering the risk conferred by genetic factors (SNPs).

Firstly we considered the different frequencies of HLA genotypes in probands and controls and secondly the non-HLA SNPs genotypes frequencies. According to the Bayesian approach, in each run we used the score obtained in the previous step as the new *a priori* value for the following step (see Text S1). By this approach, it was possible to assign to each combination of HLA plus LPP, RGS1 and REL SNPs genotypes a BS. Secondly, we assigned to each sib (validation set) a BS in dependence of the specific HLA+SNPs genotype combination found. We arbitrarily established the median BS as discrimination threshold between low and high risk sibs and evaluated how affected and unaffected sibs were distributed. We performed G statistic by using the R "ROCR" statistic package to strength the interpretation of the results.

In order to produce a more realistic Bayesian Risk Score, we then considered the mean HLA related risk [1] as the *a priori* risk

(instead of 0.5) to be fit in the first step of the Bayesian equation. The corresponding scores (higher or lower than the median) were standardized by the starting HLA related risk.

## Supporting Information

**Table S1** Allelic frequencies observed in Training set (Trios) and in Validation set (sibs). (DOC)

**Table S2** Non-HLA genotypic frequencies and HLA groups frequencies observed in Training set (Trios) and in Validation set (sibs). (DOC)

**Table S3** Classification according to the HLA genotype [1]. (DOC)

**Text S1** Bayesian theorem. (DOC)

## Author Contributions

Conceived and designed the experiments: LG LS. Performed the experiments: VI NT MVE AC MPS FT. Analyzed the data: VI MP SC LG. Contributed reagents/materials/analysis tools: LG LS SC. Wrote the paper: LG LS VIMP NT.

## References

1. Bougey M, Calzavara G, Tinto N, Gensmer D, Marguerite-Jeanin P, et al. (2007) HLA related genetic risk for celiac disease. *Gut* 56: 1054–1059.
2. Bonamico M, Ferri M, Mariani P, Nenna R, Thamsi E, et al. (2006) Serologic and genetic markers of celiac disease: a sequential study in the screening of first degree relatives. *J Pediatr Gastroenterol Nutr* 42: 150–154.
3. Mustalhti K, Csatari C, Reunanen A, Fikri E, Heier M, et al. (2010) The prevalence of celiac disease in Europe: results of a centralized, international mass screening project. *Ann Med* 42: 387–395.
4. Goto T, Romano R, Goto T, Ikeda C, Nishida S, et al. (2002) The first large population based twin study of celiac disease. *Gut* 50: 634–638.
5. Schuppan D, Jucker V, Beissert D (2009) Celiac disease: from pathogenesis to novel therapies. *Gastroenterology* 137: 1912–1923.
6. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, et al. (2007) A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 39: 827–829.
7. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, et al. (2008) Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 40: 395–402.
8. Romanos J, Beissert D, Trynka G, Zhernakova A, Bartels MT, et al. (2009) Six new celiac disease loci replicated in an Italian population confirm association with celiac disease. *J Med Genet* 46: 60–63.
9. Dima R, Martinez A, Remon-Aguero M, Malameda C, Blanco I, et al. (2009) Association of IL18RAP and OGR3 with celiac disease in the Spanish population. *J Med Genet* 46: 617–619.
10. Garner CP, Murray JA, Ding YC, Tim Z, van Heel DA, et al. (2009) Replication of celiac disease UK genome-wide association study results in a US population. *Hum Mol Genet* 18: 4219–4225.
11. Trynka G, Zhernakova A, Romanos J, Franke L, Hunt KA, et al. (2009) Celiac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. *Gut* 58: 1078–1083.
12. Amundsen SS, Rundberg J, Adenovic S, Gadjimolir AH, Acher H, et al. (2010) Four novel celiac disease regions implicated in an association study of a Swedish-Norwegian family cohort. *Genet Immunol* 11: 79–86.
13. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, et al. (2010) Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 42: 295–302.
14. Romanos J, van Duijn OC, Nohr IM, Trynka G, Zhernakova A, et al. (2009) Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. *Gastroenterology* 137: 834–840.
15. Liegert J, Demaraju S, Paulin B, Cook I, Dufour J, et al. (2004) Predictive models for breast cancer susceptibility from multiple single nucleotide polymorphisms. *Clin Cancer Res* 10: 2725–2737.
16. Schaid DJ (1996) General score tests for associations of genetic markers with disease using cases and their parents. *Genet Epidemiol* 13: 423–440.
17. Pennington DJ, Silva-Santos B, Shima J, Theodoridis E, Polla C, et al. (2003) The inter-relatedness and interdependence of mouse T cell receptor gamma-delta+ and alpha-beta+ cells. *Nat Immunol* 4: 991–998.
18. Son JS, Sahoo A, Chae CS, Hwang JS, Park ZY, et al. (2011) JunB and c-Rel cooperatively enhance Foxp3 expression during induced regulatory T cell differentiation. *Biochem Biophys Res Commun* 407: 141–147.
19. Liu H-C, Smith KA (2011) The roles of c-rel and interleukin-2 in tolerance: a molecular explanation of self-nonself discrimination. *Immunol Cell Biol* 89: 27–32.
20. Abadie V, Solid LM, Barcain LB, Jahn B (2010) Integration of Genetic and Immunological Insights into a Model of Celiac Disease Pathogenesis. *Annu Rev Immunol* 28: 493–525.
21. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition (1990) Revised criteria for diagnosis of celiac disease. *Arch Dis Child* 65: 909–911.
22. Thomson G (1993) Mapping disease genes: family-based association studies. *Am J Hum Genet* 57: 487–498.